EXHIBIT A

UNITED STATES DISTRICT COURT

FOR THE NORTHERN DISTRICT OF OKLAHOMA

STATE OF OKLAHOMA, ex. rel. W.A. DREW EDMONDSON, in his capacity as ATTORNEY GENERAL OF THE STATE OF OKLAHOMA and OKLAHOMA SECRETARY OF THE ENVIRONMENT, J.D. Strong, in his capacity as the TRUSTEE FOR NATURAL RESOURCSE FOR THE STATE OF OKLAHOMA Plaintiffs, v. TYSON FOODS, INC., TYSON POULTRY, INC., TYSON CHICKEN, INC., COBB-VANTRESS, INC., AVIAGEN, INC., CAL-MAINE FOODS, INC., CAL-MAINE FARMS, INC., CARGILL, INC., CARGILL TURKEY PRODUCTION, LLC, GEORGE'S INC., GEORGE'S FARMS, INC., PETERSON FARMS, INC., SIMMONS FOODS INC., and WILLOW BROOK FOODS, INC., Defendants.	)	Case No. 05-CV-329-GKF-PJC

EXPERT REPORT OF

William H. Desvousges, Ph.D.

Gordon C. Rausser, Ph.D.

INTRODUCTION AND SUMMARY OF OPINIONS

This report provides the expert opinions of Dr. William H. Desvousges and Dr. Gordon C. Rausser in response to the plaintiffs' estimate of damages for reducing alleged aesthetic effects and ecosystem injuries from phosphorous in the Illinois River System and Tenkiller Lake to levels that purportedly existed in the 1960s. The plaintiffs in this case hired Stratus Consulting (Stratus) to prepare the damage estimates. Stratus prepared two natural resource damage assessment (NRDA) reports entitled: *Natural Resource Damages Associated with Aesthetic and Ecosystem Injuries to Oklahoma's Illinois River System and Tenkiller Lake*[1] and *Natural Resource Damages Associated with Past Aesthetic and Ecosystem Injuries to Oklahoma's Illinois River System and Tenkiller Lake*.[2]

The first opinion in this report addresses the recreational use of Tenkiller Lake and the Illinois River. Our analysis is based on several data sources, including data collected, but not used, in the Stratus expert reports. As our analysis below shows, Stratus first conducted a study that interviewed users of Tenkiller Lake and the Illinois River during several months in 2006 (hereafter, user intercept survey). Subsequently, Stratus conducted a telephone survey of a random sample of Oklahoma residents that included both users and people who do not use the resources to gauge their knowledge and awareness of water quality issues in both the Illinois River System and Tenkiller Lake (hereafter, telephone survey). Thus, Stratus could have conducted a damage assessment based on people's actual recreation decisions and determined whether those decisions were affected by water quality. Instead, they chose to rely solely on the hypothetical CV survey, in part because people's perceptions of water quality for the Illinois River System and Tenkiller Lake, based on both the intercept and telephone surveys, were more favorable than the plaintiffs' experts would have preferred.

As part of our investigation, we collected data on recreational uses over time in the Illinois River System and Tenkiller Lake from public data sources. Our analysis shows that the recreation use data from both Stratus and public sources indicate that

[1] The authors of this report include: David J. Chapman, Richard C. Bishop, W. Michael Hanemann, Barbara J. Kanninen, Jon A. Krosnick, Edward R. Morey, and Roger Tourangeau.
[2] The authors of this report include: W. Michael Hanemann, David J. Chapman, and Richard C. Bishop.

the Illinois River and Tenkiller Lake are valuable recreation resources and that users find them to be clean and enjoyable. Thus, there has been no lost benefit to the public as a result of an alleged injury to water quality. Consequently, there are no natural resource damages from reduced recreational use. Details supporting this opinion appear in Section 2 of this report.

To further investigate whether alleged injuries to water quality in Tenkiller Lake has impacted residents of the area, we collected and analyzed data on property transactions for single family residences near the lake during an fourteen year period, from 1995 to 2008. Our analysis compares transaction prices for this period to those for neighboring properties. We find no evidence that alleged water quality injuries have reduced property values near Tenkiller Lake. For an additional benchmark, we compared data on sales transactions near portions of Lake Eufaula over the same period of time. Again, we find no evidence that the alleged injuries have affected sales prices of properties near Tenkiller Lake. Section 3 of this report provides the details supporting these opinions.

Additionally, we offer opinions that directly respond to the plaintiffs' estimates of past and future damages that purportedly reflect society's total values for the natural resources at issue. Total values embody both the value that people place on these resources based on their direct experience using the resources (use values) and values that are independent of their use of the resource (nonuse values).

The plaintiffs' estimate of future damages depends solely on the use of the contingent valuation (CV) methodology and fails to include any of the data from the first two Stratus surveys discussed above, or any other public sources. Based on our evaluation, the Stratus damages studies do not conform to the 43 CFR Part 11, the U.S. Department of the Interior natural resource damage assessment regulations. The report was not preceded by a preassessment screen or an assessment plan. Additionally, the report was not available for public review and comment before submittal to the court. Moreover, as we discuss in detail below, the Stratus CV study and the past damages benefits transfer do not meet the requisite reliability conditions that are required to comply with the 43 CFR Part 11 regulations.

Two features of the plaintiffs' CV methodology are especially critical in regards to evaluating its validity. First, the CV methodology is based on respondents' answers to a hypothetical survey questionnaire. Second, CV cannot reliably measure nonuse values, a component of total values. Because no behavior can be observed with respect to nonuse values, they cannot be directly measured or validated by using objective criteria. The inclusion of nonuse values in the plaintiffs' damages estimates means that it is not possible to externally validate the respondents' answers. Concerns about CV's ability to reliably measure nonuse values prompted the National Oceanic and Atmospheric Administration (NOAA) to form a Blue Ribbon Panel to consider this question. This panel produced a set of guidelines for conducting CV studies (Arrow, et al. 1993)[3], which are discussed in detail in Section 3, 4, and 5 of this report. These guidelines comprise the only set of detailed guidance for conducting CV studies. As such, they apply to damage assessments conducted under both the NOAA regulations (15CFR990) and the Department of Interior (43 CFR Part 11) regulations.

The evaluation of validity involves a careful review of the Stratus CV survey questionnaire as it relates to the accuracy, balance, and other factors that may lead to potential biases in the survey respondents' answers. Additionally, the evaluation requires a careful examination of the consistency of the respondents' answers with basic economic principles and resulting validity tests that have evolved in the peer-reviewed literature.

The results of this validity assessment of the Stratus CV methodology are directly relevant to the estimate of past damages, which relies on the CV estimates as a starting point to measure past damages. Specifically, the Stratus report purportedly relies on a methodology known as benefits transfer, in which the results of one study are used to estimate the benefits for another resource. In the novel Stratus application of benefits transfer, the current damage estimate for the Illinois River System and Tenkiller Lake are "transferred" to estimate the past damages. In the Stratus report, past damages are estimated by assuming that the current CV estimates can be extrapolated into the past using some simplistic assumptions about the relationship

[3] Arrow, K., R. Solow, P.R. Portnoy, E.E. Leamer, R. Radner, and H. Schuman. 1993. "Report of the NOAA Panel on Contingent Valuation." 58 *Fed. Reg.* 4601 *et. seq.* January 15.

between current and past injuries to water quality in the Illinois River System and Tenkiller Lake. These past damages are then augmented by adding compound interest.

Based on our review of the Stratus reports and the CV study contained therein, we have concluded that it is an unreliable basis for measuring the potential aesthetic and ecosystem damages to the Illinois River System and Tenkiller Lake from increased phosphorous loads. The CV methodology as applied in this case contains substantial biases all of which render it unreliable for estimating potential damages. Specifically, we conclude that:

- The Stratus survey questionnaire contains biased, misleading and factually incorrect information. Thus, the Stratus questionnaire fails to meet the NOAA panel guidelines for a neutral, balanced presentation of information.
- Many respondents were expressing values for resources different than those intended by the survey designers. Our analysis shows that approximately 80 percent of respondents rejected at least one key feature of the CV survey. In fact, more than 40 percent of respondents indicated that they thought the hypothetical restoration program would benefit other lakes and rivers in Oklahoma. This result, which is often found in CV surveys like the one conducted by the plaintiffs, renders their answers meaningless as a basis for valuing changes in water quality in the Illinois River System and Tenkiller Lake.
- The Stratus survey contains substantial amounts of hypothetical bias (the difference between what people actually do and what they say they would do), a phenomenon that has been demonstrated repeatedly in the literature on CV. In fact, more than 30 percent of respondents in the Stratus statistical analysis paid no state income tax, which was the method in which the hypothetical payments were expressed. Such bias invalidates the CV results as a basis for measuring damages.
- The format used in the Stratus CV survey, in which people express votes on a hypothetical restoration project in a simulated referendum, does not mitigate the hypothetical bias in the survey. We demonstrate substantial differences between a real referendum and the contrived one described in the Stratus report.
- The Stratus report fails in its attempts to correct for hypothetical bias.
- The Stratus damage estimates do not correspond to basic economic principles of demand. Specifically, the Stratus valuation responses do not conform to the law of demand or to fundamental principles related to the responsiveness to income.

- The scope test devised by Stratus, which measures whether respondents' answers are sensitive to changes in the extent of the injury, contains a fatal flaw that prevents it from serving as a test of potential validity. Their test confounds the amount of injury with the effectiveness of the purported restoration project. Moreover, the outcome of the scope test is largely an artifact of the large sample size the analysts chose for the base survey version.

- The Stratus survey contains nonresponse bias, which results from unknown differences between survey respondents and the people who failed to respond to the survey. Given that the Stratus survey was unable to reach at least 45 percent of potential respondents, this bias is significant. The plaintiffs' attempts to adjust for this bias are ineffective.

- The Stratus damage estimates include an unsubstantiated and flawed assumption that most households in Oklahoma would have the same value for aesthetic and ecosystem changes that were estimated in the CV survey. Our analysis shows that Stratus knows nothing about the awareness and knowledge of respondents that they did not interview, which is a critical requirement for estimating damages for a population as a whole.

- Empirical analysis of the CV results demonstrates the consistently upward bias of the Stratus analytical techniques. In other words, the CV results consistently lead to overstated estimates of damages. The empirical analysis demonstrates that the damage estimates are highly dependent on the analysts' choice of the hypothetical bids.

- The Stratus survey does not conform to all of the NOAA panel guidelines.

- Because plaintiffs' past damages depend critically on the CV study, the plaintiffs' estimate of past damages reflects all of the flaws in the CV study, which render the estimate of past damages unreliable.

- Moreover, the benefits transfer performed by Stratus does not conform to established principles and practices for conducting transfers.

The remainder of this report discusses these opinions in greater detail. Specifically, Section 2 describes the recreation surveys and related data that form the basis of our opinion about recreation uses. Section 3 provides an analysis of real estate property values on and surrounding Tenkiller Lake. Section 4 provides our analysis of the CV survey report and data and the rationales underlying our opinions. Section 5 presents our critique of the novel benefits transfer that Stratus used to estimate past damages. Section 6 describes the aggregation of future damages, while Section 7 describes our evaluation of the plaintiffs' total damages calculations.

2. RECREATION USE ANALYSIS

To evaluate potential recreation losses from increased phosphorous loadings into the Illinois River System and Tenkiller Lake, we have evaluated information based on people's actual use of their resource, and their unaided perceptions of the quality of those resources. As indicated above, we rely on the results of the Stratus user intercept and telephone survey to form our opinions about whether recreation use has been impacted by phosphorous loadings. We also rely on data from the United States Army Corps of Engineers (COE), which manages Tenkiller Lake and maintains data on recreational uses on that lake, as well as other lakes that the COE manages in Oklahoma and other states. The results of the Stratus user intercept and telephone surveys and the analysis of the COE data indicate that recreational use is increasing in the Illinois River System and Tenkiller Lake and that few residents perceive any reduction in quality of the resources and have experienced no loss in benefits from the resources. Moreover, respondents who had negative impressions of the resources more frequently mentioned other issues such as trash and unruly users. The data and interpretations are discussed below.

2.1 The Stratus user intercept survey shows users experienced high quality recreation in both the Illinois River and Tenkiller Lake.

The primary purpose of the Stratus Illinois River and Tenkiller Lake user intercept survey was to obtain current estimates of recreational use on Tenkiller Lake and the Illinois River (Stratus 2008, p. B-1). The study took place from Memorial Day weekend, 2006 to September 17, 2006. Based on the counts of cars, the number of people in the cars, length of stay, and the total number of times each of these sampling units could have been observed, the total estimated user days over the time period is 294,243 (Stratus 2008).

As part of the survey, Stratus interviewers intercepted recreators and collected information about their behavior and attitudes. Three hundred and ninety-five individuals (90 percent of those intercepted) agreed to participate in the survey. Respondents were asked: "Thinking specifically about the Illinois River/Tenkiller Lake, are there one or two things that you particularly like or dislike about recreation here?"

Seventy-eight percent of the responses were positive, and only 22 percent were negative. As shown below, negative responses range from congestion at facilities, to trash, to inadequate restroom facilities. Only three respondents, or less than one percent, mention any water quality problems related to increased algae from phosphorous.

Of all the respondents, 272 were intercepted at Tenkiller Lake and asked several questions about the quality of their experiences. When asked what they particularly liked and disliked about recreation there, 422 positive responses and 68 negative responses were given. Thirty-four percent of the respondents explicitly mentioned good water quality as something they specifically liked about Tenkiller Lake. Good water quality was second in frequency of response only to "the natural beauty and aesthetics" of the lake as a quality they particularly liked about recreation at Tenkiller Lake (45 percent).

In contrast, the most common dislike about recreation at Tenkiller Lake was trash and debris at the site (8 percent). The second most common dislike was the limited access and facilities (6 percent). And, the third most common was unruly users of the site (5 percent). Only eight, or about 3 percent, of respondents mentioned poor water quality, and when probed, none mentioned water clarity and three mentioned something related to chicken farms. Of these three, one specifically mentioned that good water quality was something they specifically liked about the lake.

Stratus intercepted 123 respondents at sites along the Illinois River. When respondents were asked what they particularly liked or disliked about recreation at the Illinois River, 223 of the responses (73 percent) were positive and 81 of the responses (27 percent) were negative (respondents gave multiple responses). The most common positive response was the convenience of the River, "it is close to home" (50 percent). The three other most common responses referred to the natural beauty of the river (43 percent), accessibility (40 percent), and lack of congestion (26 percent). Ten percent of respondents specifically mentioned good water quality as something they liked about recreation at the Illinois River.

The most frequently mentioned dislike about the site was the presence of trash and debris (24 percent). The second most common negative remark was that they did not like the limited access or facilities (14 percent). The third most common dislike was unruly users (11 percent). Finally, nine percent of the respondents mentioned congestion as a dislike. A total of six respondents (1.2 percent) mentioned poor water quality as something they disliked. However, when probed, 4 out of the 6 respondents indicated that they were referring to debris in the water not water clarity. Of the other two respondents, one mentioned algae coming from Lake Francis and the other mentioned nonspecific pollution.

The results of a 2006 Stratus user intercept survey, which are reinforced by the results of the telephone survey discussed below, show that users of the resources have a very good impression of the quality of recreation at the site and the beauty of the site. Specifically, users mention very few negative characteristics of the site, and of those, poor water quality is rarely mentioned. In fact, good water quality and scenic beauty seem to be the most commonly noted characteristics of the sites. Therefore, the user intercept survey indicates that recreators have a positive impression of water quality and enjoyed high quality experiences on both Tenkiller Lake and on trips to the Illinois River.[4] The results of this study, based on actual users during their recreation trips, are in stark contrast to the results of the Stratus CV study, which is based on responses to hypothetical questions, and was conducted only after dosing respondents with biased and misleading information about water quality conditions in the two waterbodies.

[4] According to one of the consultants hired by the State, "Water clarity does not seem to be a big issue among floaters. More concerned with litter, crowds and cost of floating experience? Do people care about the P loadings? Do they know about them and their affects (*sic*)?"(Breffle 2004).

2.2 The Stratus telephone survey reinforces the results of the user intercept study that recreational use is not impacted by water quality.

Stratus also conducted a telephone survey in 2006 to identify respondents' knowledge and use of Oklahoma's waterbodies (particularly Tenkiller Lake and the Illinois River), to determine respondents' awareness and perceptions of water quality problems, and to evaluate the impact of media attention on the poultry industry (Stratus 2007). They used a stratified random sample of the entire state of Oklahoma, which over-sampled residents who live closer to the waterbodies. The results include 400 completed surveys.

As stated by Stratus in its report on the telephone survey, "a critical component of this survey is to determine whether respondents know of any water quality problems in Oklahoma and what they perceived to be the causes of these issues" (Stratus 2007, p. 4). Questions 16 and 17 were designed to elicit the respondents' impressions about the sites without prompting them with any suggestion of water quality as an issue (Stratus 2007). Stratus indicates that this approach "allows respondents to state what they remembered without influencing their views by specifically mentioning water quality in the question" (Stratus 2007, p. 4).

2.2.1 The telephone survey respondents have a favorable impression of the Illinois River and Tenkiller Lake.

Questions 16 and 17 state: "What impression do you have of the Illinois River or Tenkiller Lake? Is there anything especially good or bad about it?" Again, in contrast to the Stratus CV study, the responses are open-ended and provide impressions that are not tainted by the interviewer's suggestions or any leading text. Research has shown that responses to survey questions are heavily influenced by the information that the questionnaire/interviewer provides (Payne, Bettman, and Schkade 1999; Bettman, Luce and Payne 1998; Frederick and Fischhoff 1998; and Rea and Parker 2005).

Considering only those respondents who had visited the Illinois River, most of the responses given to question 16 were positive (81 percent). Some of the impressions provided by the respondents are listed below:

- "I like it because it has clear drinking water and a strong current to canoe on it. I learned to swim there because it held me up."
- "We need to make sure it stays the way it is – clean and clear and we need to preserve it and we need to promote catch and release."
- "Spring-fed clear."
- "That's my favorite river. Really pretty. It's a great river – not particularly deep – you can see the fish it's that clear – big."
- "Clear and cold – trout fishing – canoeing."
- "Clear water."
- "It's just good water."
- "Oh it's a beautiful float river and it's good for fishing"
- "Fun place for recreation"
- "It's clean and the people in Arkansas need to keep their chicken litter in Arkansas and out of the river."
- "Very pretty"
- "I love it very clear it's my favorite river and we float every year."
- "The scenery, it's very nice"
- "Clarity of the water."
- "Absolutely beautiful"
- "It's very scenic, very good"
- "Best place I have ever vacationed"
- "It's good for activities"
- "I know it's real cold and it's good for rafting"
- "Good river to go canoeing"
- "It's beautiful and the water is clear"
- "Its water is clear it's good for boating"
- "Pretty. Family oriented."
- "The good is you get to float in it. I don't know anything bad about it."
- "I like the scenery it had beautiful water it's a great place to take a family on a family outing."

There were 233 respondents who had not visited the Illinois River. One hundred thirty-one of those gave some impressions of the river. Most of the impressions were positive or were neutral, such as "heard people raft there." There were 103 of these positive/neutral responses. Twenty-three of these respondents specifically mentioned the beauty of the river and/or the clear, clean water. Of the 28 negative impressions, half mentioned that they had heard there was a problem with poultry litter (6 percent).

Of the respondents that had visited Tenkiller Lake, 100 provided their impression of the lake, in response to Question 17 on the survey. The vast majority of the impressions were good (92 percent). Most of the impressions referred to the scenic beauty, clear water and recreation opportunities. Some of the impressions given by respondents that describe the beauty and clarity of the water are listed below:

- "It was a pretty, nice lake"
- "Just the beauty of the water"
- "I love it. I've got land on Tenkiller"
- "Just pretty water"
- "Like the trees…the rocks and its clean…it has the clear water"
- "It's pretty"
- "It has nice clear water"
- "Cleanest lake in Oklahoma"
- "It's a beautiful place"
- "It has clear water – clearer than the others"
- "It's beautiful –deep and cold"
- "Was the one you can see your feet in the water, it is really clear and you can see fish"
- "Beautiful a lot of people go there"
- "Just know that it's clearer than most"
- "It's clear and it's beautiful"
- "Clear water – the Illinois River runs into the Tenkiller"

- "I like to go there and boat, it's so pretty"
- "The water is clear it's beautiful and it's not over-crowded"
- "Prettiest lake in Oklahoma. Everything is good about Tenkiller"
- "It's pretty and blue, not dirty"
- "The only thing I remember it it's a clear lake, they SCUBA dive down there"
- "Clear water and scenic. Nice facilities. Great fishing."
- "Pretty lake – cold, deep water"
- "The water is very clear and it has beautiful cliffs …"
- "Its scenic and the water is very clear and very deep"
- "It's really clear and really cool to SCUBA dive…"
- "It's very scenic nice water"
- "Nicest lake in the state"
- "It was very scenic and water very clear"
- "It's very deep and clear – with the drought water stayed up- good for SCUBA"
- "Nice to dive in – SCUBA diving in the clear water"
- "Great lake with beautiful blue water"
- "It's a beautiful lake"
- "The cleanliness and the way it is laid out are great"
- "Camp site is nice and the water is good"
- "…one of the cleanest lakes in the state right now"
- "I thought it was very clear"
- "Lake is great – water is clear.."
- "Supposed to have real clear water"
- "Beautiful, I don't know anything bad about it"
- "Real pretty – clean"

There were only eight negative impressions given as responses to Question 17 by visitors of Tenkiller Lake. Of those eight, there were three respondents who mentioned diminishing clarity and one that mentioned poultry litter.

There were 207 responses to Question 17 by respondents who had not visited Tenkiller Lake. Only 88 of these respondents had an impression of the lake. Positive impressions were by far the most common, representing 95 percent of the total, or 84 responses. There were four negative impressions given of the lake. One of these responses referred to pollution and one to reduced water clarity.

The responses to Questions 16 and 17 indicate that, without prompting the respondents about water quality being an issue, users' impressions were very positive. The overall impression of the Illinois River is that it is a great recreation area and very scenic. Tenkiller Lake is described by multiple respondents as the most beautiful lake in Oklahoma. The dominant feature of the lake, in the minds of the respondents, seems to be its clear water. There is no indication that recreational users of the Illinois River and Tenkiller Lake are experiencing any lost benefits in the use of these resources as a result of a reduction in water quality or other alleged ecosystem impacts. Even nonusers seemed unaware of any water quality issues associated with the Illinois River and Tenkiller Lake. The very low level of awareness about "poultry litter" was usually described as something they "heard about."

2.2.2 Prompting biases people's impressions of the sites.

After eliciting uncorrupted impressions about the site, the respondents were prompted to respond to the question, "Have you heard of any issue or concerns relating to the Illinois River or Tenkiller Lake or are you unaware of any issues or concerns there?" Forty percent respond that they have heard of issues or concerns, even though less than 20 percent of respondents overall had mentioned any negative impressions of the Illinois River and Tenkiller Lake before being prompted. When asked to describe the issue or concern, Q19, 41 percent of those that were aware of a problem mentioned poultry, whereas before being prompted less than three percent mentioned poultry. These results indicate that even respondents who were aware of the alleged poultry

litter problem did not consider it significant enough to mention as one of their impressions of the lake or river.

Next, the interviewers probed the respondents that did not mention water quality issues as a concern of which they were aware, by asking them specifically if they were aware of, or had heard of, water quality problems on Tenkiller Lake or the Illinois River. As a result of this explicit prompting, 79 respondents who were not aware of problems before became aware of water quality problems when prompted. Again, this result provides evidence that prompting a respondent can produce different results. A respondent does not want to appear unknowledgeable to an interviewer and will often agree with the interviewer when prompted to do so. The results of the survey then exhibit a bias in the statement of opinion. This phenomenon is well-established in the survey literature (Leggett, et al. 2003; Tourangeau, Rips, and Rasinski 2000).[5]

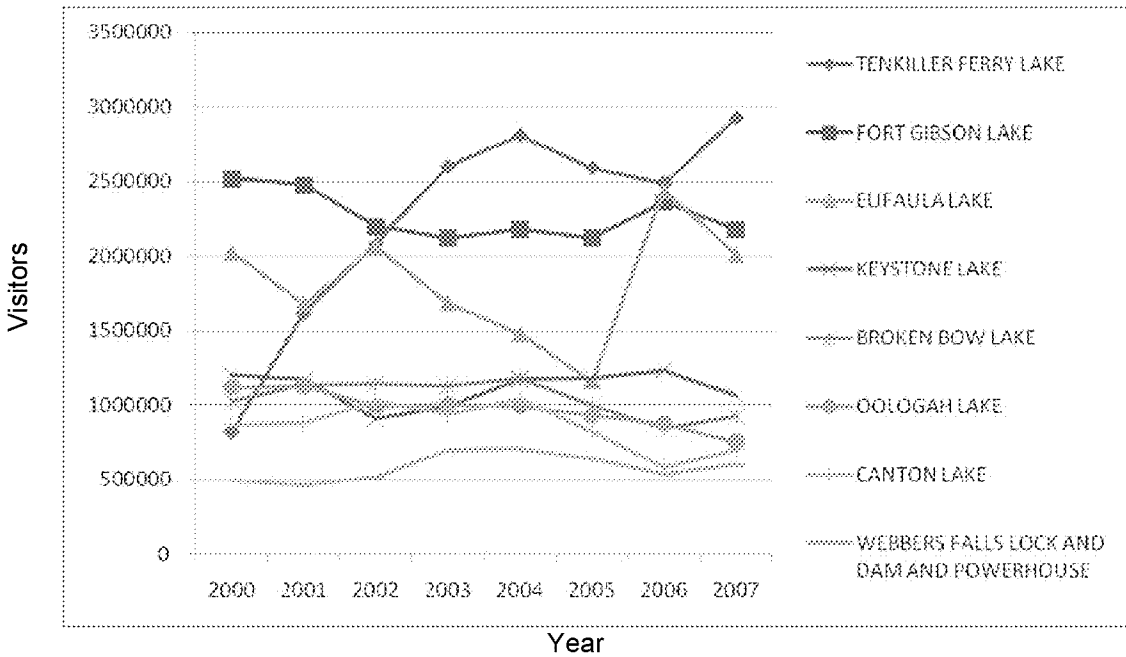
In summary, the Stratus telephone survey provides similar results to the user intercept survey. Respondents have favorable impressions of both the Illinois River and Tenkiller Lake, with good water quality being one of the desirable features mentioned most often. The Stratus telephone survey is important because it is based on a statewide survey of Oklahoma residents who were not dosed with biased information about water quality.

2.3 Data on user fees and from the COE show that Tenkiller Lake and the Illinois River are vibrant recreation resources.

Tenkiller Lake is a recreation site managed by the COE. There are more than 400 COE sites nationally, 28 in Oklahoma. The COE maintains a visitation record for all of its recreation sites. In 2007, Tenkiller Lake was the most visited COE lake in Oklahoma and in the top 10 percent nationally. The trend in visitation for the top eight most visited COE sites in Oklahoma has been fairly stable with exception of Tenkiller Lake. Tenkiller Lake has seen a rise in visitation from 818,522 in 2000 to 2,924,047 in 2007, nearly a 300 percent increase. The remaining top eight sites have seen either a decrease in visitation or have remained unchanged since 2000 (Figure 2.1).

[5] It also is endemic to the Stratus CV survey, which is discussed in Section 4.1 below.

Figure 2.1: Number of visitors to the eight most popular U.S.Corps of Engineers lakes in Oklahoma, 2000-2007



In addition, Tenkiller Lake is a popular site for bass tournaments. Annual tournament results showed that Tenkiller Lake consistently was ranked in top 20 lakes included in the results (ODWC 2001; ODWC 2002a; ODWC 2003; ODWC 2004; ODWC 2007).[6] Field and Stream Magazine (2008) recently cited Tahlequah as one of best fishing towns in America. As part of the rationale for the selection, the magazine praises the largemouth bass fishing in Tenkiller Lake. Specifically, the magazine notes that:

> Lake Tenkiller is a gem. Its waters are remarkably clear, and its 130 mile shoreline is picturesque (p.1-2).

Similar statements can be found in the 2009 Lake Tenkiller Visitors Guide:

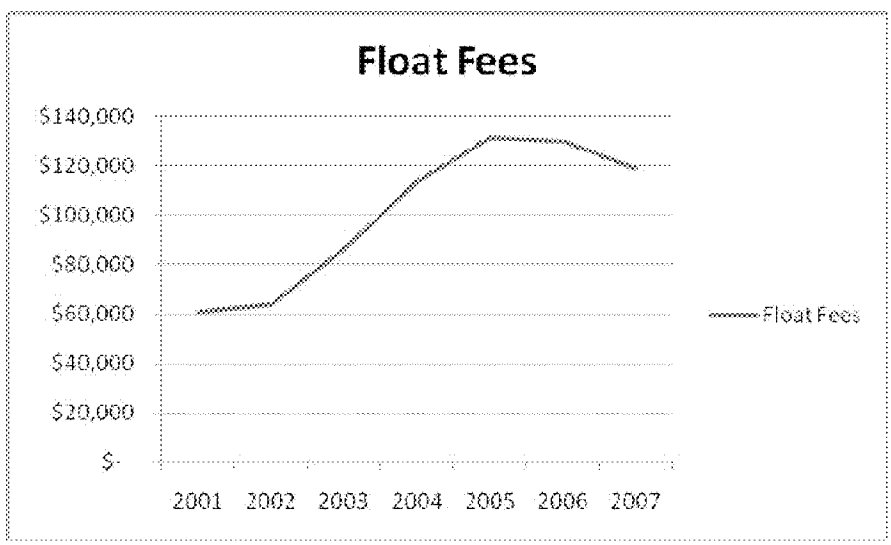
[6] In 2000, Tenkiller Lake experienced the only documented fish kill resulting from Largemouth Bass Virus in Oklahoma. Electrofishing catch rates of bass in 2002 showed the lowest numbers since 1990 (ODWCb 2002). The Largemouth Bass Virus may have affected the outcome of Bass Fishing Tournaments for several years following the fish kill.

> On the main body of the lake visitors can experience the best water clarity in the State. Each year, countless scuba enthusiasts travel to Tenkiller to dive in its deep clear waters (p.5).

The Visitors Guide characterizes Tenkiller Lake as the Emerald jewel in Oklahoma's Crown of Lakes (Greater Tenkiller Area Association 2009).

Similarly, the Illinois River is a vibrant recreation resource that shows trends of increasing use. For example, the state collects a fee of $1 per person to float the Illinois River. Use of the river can be tracked through the collection of these fees. As indicated by the yearly fees collected (Figure 2.2), use of the Illinois River by rafters increased steadily from 2002 to 2005 when total user fees collected peaked at more than $130,000. Total fees collected dropped slightly in 2007, but still remained at around $120,000.

Figure 2.2: Revenues from Float Fees on the Illinois River



The COE visitation results and the float fee revenues indicate that there has not been a reduction in visitation to Tenkiller Lake or the Illinois River as the result of any alleged injury. In fact, Tenkiller Lake visitation has increased substantially during the past several years. These visitation records support our opinion that recreation users

have not experienced a reduction in the quality or quantity of their recreation experiences over the past eight years. Moreover, these results are reinforced by those in both the Stratus user intercept survey and the telephone survey of Oklahoma residents, as well as supplemental information on fishing including a national publication touting fishing on Tenkiller Lake. Thus, there is no basis to estimate potential damages related to changes in water quality for Tenkiller Lake and the Illinois River.

In order to understand the factors that affect visitation to COE lakes in Oklahoma, we developed a regression model using lake characteristics. Characteristics describing water quality, facilities, lake characteristics, and distance from population centers are used to explain the level of visitation. Specifically, we use the 22 COE lakes in Oklahoma that have data on lake levels. Measurements of water clarity were obtained from the Oklahoma Water Resources Board. Lake characteristic data were obtained from the COE Tulsa district website. Visitation data were provided by the COE Institute for Water Resources for the years 2000 to 2007.

Our analysis considered a variety of lake characteristics, including the water level in the lake. In particular, water levels deviating from normal would likely affect recreation. In addition, the size of the lake could affect recreation and aesthetics, which would impact visitation. To reflect lake size in the model, we include variables describing the ratio of shoreline to lake acres and lake depth. Facilities available at the site would impact recreation. Therefore, we include variables identifying the number of campsites, boat ramps, and a qualitative variable indicating that the lake had at least one state park. All the facilities data were available on the COE Tulsa district website and the Oklahoma Tourism website.[7]

[7] http://www.swt.usace.army.mil/; http://www.touroklahoma.com/

Table 2.1: Variables Included in the Model

Variable names	Description
meanclarity	The average water clarity measurement in a lake for the year (cm)
lakelevel	The average deviation from normal in the months of June, July and August (feet)
campsites	The number of campsites on the lake
boatramps	The number of boat ramps on the lake
shoreacres	The ratio of shoreline miles to lake acres (miles/acre)
statepark	0,1 indicating the presence of a state park on the lake.
distance	Distance from the closest metropolitan area (Tulsa, Oklahoma City, Dallas) (miles)
lakedepth	Normal water elevation as indicated by the Corp of Engineers (feet)

As shown in Table 2.2, the results indicate that the regression is about to explain about two-thirds of variation in visitation levels. The model also shows that above normal lake levels lead to decreased visitation. As one would expect, lakes with more facilities, such as campsites and state parks, have higher levels of visitation. Lakes with a higher ratio of shoreline to lake acres receive fewer visitors. These would be narrower lakes with many bays and inlets. The number of boat ramps also significantly influences the level of visitation.

To evaluate the potential effect of water quality on visitation at COE lakes, we used the average water clarity of the lake. In other specifications, we used the minimum and maximum water clarity measurements for the season. Our analysis indicates that none of the indicators for water clarity were found to significantly predict visitation. Thus, aggregate visitation for the COE sites for the years 2000 to 2007 was not impacted by variation in water quality, as measured by water clarity levels. The model results also show that there is no significant time trend in visitation across the sites. These results provide further support for our conclusion that recreation at Tenkiller Lake has not been impacted by changes in water quality and that recreators have not experienced any potential losses from alleged injuries attributable to increased phosphorous loadings from the application of poultry litter.

Table 2.2: Recreation Model Results

Invisits	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
meanclarity	.0029796	.0019299	1.54	0.125	-.0008351	.0067942
lakelevel	-.0475354	.0235078	-2.02	0.045	-.0940003	-.0010706
campsites	.0034361	.0007994	4.30	0.000	.0018559	.0050163
boatramps	.1165248	.0303536	3.84	0.000	.0565287	.1765209
shoreacres	-24.80669	6.60842	-3.75	0.000	-37.86873	-11.74465
lndistance	-.3431398	.1019084	-3.37	0.001	-.5445693	-.1417103
statepark	.2224358	.1223296	1.82	0.071	-.0193577	.4642294
lakedepth	-.0002408	.0001781	-1.35	0.178	-.0005927	.0001112
y2001	-.0227072	.1968977	-0.12	0.908	-.4118902	.3664758
y2002	.0569172	.195921	0.29	0.772	-.3303353	.4441697
y2003	.0123243	.198466	0.06	0.951	-.3799587	.4046073
y2004	.0737207	.1960291	0.38	0.707	-.3137455	.4611868
y2005	-.1082455	.1975569	-0.55	0.585	-.4987316	.2822406
y2006	-.2467322	.1997219	-1.24	0.219	-.6414976	.1480332
y2007	-.0474614	.2427033	-0.20	0.845	-.5271826	.4322598
_cons	13.26871	.6123233	21.67	0.000	12.0584	14.47901

2.4 References

Bettman, J.R., M.F. Luce, and J.W. Payne. 1998. "Constructive Consumer Processes." *Journal of Consumer Research* 25: 187-217.

Breffle. 2004. Memo. (Stratus0026104_Breffle_Ques_trip.doc)

Field and Stream. 2008. "The Best Fishing Towns in America." http://www.fieldandstream.com/photos/gallery/kentucky/2008/01/best-fishing-towns-america?photo=15#

Frederick, S. and B. Fischhoff. 1998. "Scope (in)sensitivity in Elicited Valuations." *Risk, Decision, and Policy* 3: 109-123.

Greater Tenkiller Area Association. 2009. Lake Tenkiller 2009 Visitor's Guide.

Leggett, Christopher G., Naomi S. Kleckner, Kevin J. Boyle, John W. Duffield, and Robert Cameron Mitchell. 2003. "Social Desirability Bias in Contingent Valuation Surveys Administered Through In-Person Interviews." *Land Economics* 79(4): 561-575.

Oklahoma Department of Wildlife Conservation. 2007. "Oklahoma Bass Tournaments: Annual Report 2007."

Oklahoma Department of Wildlife Conservation. 2004. "Oklahoma Bass Tournaments: Annual Report 2004."

Oklahoma Department of Wildlife Conservation. 2003. "Oklahoma Bass Tournaments: Annual Report 2003."

Oklahoma Department of Wildlife Conservation. 2002a. "Oklahoma Bass Tournaments: Annual Report 2002."

Oklahoma Department of Wildlife Conservation. 2002b. "Largemouth Bass Electrofishing Results 2002." http://www.wildlifedepartment.com/electrofishing2002.htm

Oklahoma Department of Wildlife Conservation. 2001. "Oklahoma Bass Tournaments: Annual Report 2001."

Payne, J.W., J.R. Bettman, and D.A. Schkade. 1999. "Measuring Constructed Preferences: Towards a Building Code." *Journal of Risk and Uncertainty* 19: 243-270.

Rea, Louis M. and Richard A. Parker. 2005. *Designing and Conducting Survey Research: A Comprehensive Guide*. San Francisco: Josey-Bass Publishers.

Stratus 2008. "Short Report for Lake Tenkiller and Illinois River Recreational Use (LT/IRRU) study – Draft." (Stratus0026231_LTIR_RU_short_for_main_report.doc)

Stratus 2007. "Oklahoma Watershed Short Telephone Survey Report (Draft)." March 29. (Stratus0027772_STS.Draft_NEW.doc)

Tourangeau, Roger, Lance J. Rips, and Kenneth Rasinski. 2000. *The Psychology of Survey Response*. Cambridge: Cambridge University Press.

3. ANALYSIS OF REAL ESTATE PROPERTY VALUES ON AND SURROUNDING TENKILLER LAKE

Real estate property values offer yet another way to objectively measure whether or not consumers perceive a decline in value of Tenkiller Lake as a result of reduced water quality or other potential aesthetic effects, such as those alleged in the Stratus CV questionnaire. As we outline here, there is no evidence that property values on and surrounding Tenkiller Lake have exhibited any decline or stagnation over a 14 year period from 1995 to 2008 during which plaintiffs claim declining water quality.

In order to test whether the value of properties located near Tenkiller Lake were affected by phosphorous, we compared it to a similar lake that is not allegedly contaminated with phosphorus: Lake Eufaula.[8] While Lake Eufaula is considerably larger than Tenkiller, both lakes are approximately the same distance from two major cities: Tulsa and Oklahoma City. Similarly, they are approximately the same distance from Tulsa International Airport. Other than scuba diving, both lakes offer similar recreational activities such as fishing, boating, camping, swimming, golfing, et cetera. As noted in the previous section, both lakes are managed by the COE. Both also support similar levels of recreation activity. The table below outlines similarities and differences between the two lakes.

Table 3.1: Similarities and Differences between Lake Tenkiller and Lake Eufaula

Lake Tenkiller	Lake Eufaula
Man-made, clear water, good for scuba-diving	Man-made
130 miles of shoreline, 12,900 surface acres	600 miles of shoreline, 102,200 surface acres. Largest lake in OK.
47.7 m from Fort Smith AK, 94.5 m from Tulsa OK, 160 m from OKC	86.4 miles from Ft. Smith AR, 88.2 miles from Tulsa, 134 m from OKC
95 miles from Tulsa International Airport	88.7 miles from Tulsa International Airport
Nearby communities: Cookson, Keys, Gore, Sallisaw, Tahlequah, Vian	Nearby communities: Checotah, Eufaula, Kiamichi, McAlester, Muskogee, Stigler, Crowder

[8] While Tenkiller Lake is on Department of Environmental Quality's 303(d) list (September 2008, Appendix C) for being aesthetically impaired due to phosphorus, no parts of Lake Eufaula are on the list for phosphorus impairment.

Other things being equal, a home located on or near a lake that is aesthetically impaired would be expected to have a lower price than a similar house located lake that is not impaired. As such, if Tenkiller Lake is aesthetically impaired, then we would expect the property values to be significantly and negatively affected by being located on or near Tenkiller Lake. Michael, Boyle and Bouchard (2000) find that decreases in water clarity as a result of eutrophication lead to reduced property values in Maine lakes. Gibbs, et al. (2002) find similar results for lakefront property in New Hampshire where decreased water clarity resulted in property value reductions ranging from .9 to 6 percent. Gibbs, et al. (2002) indicate that the implicit prices for improved water quality are comparable between the two states, despite differences in the sizes and other characteristics of lakes between the two states. Poor, Pessagno, and Paul (2007) evaluate the effects of changes in total suspended solids and dissolved inorganic nitrogen on property values in the St. Marys watershed. They found that poorer water led to reduced property values.

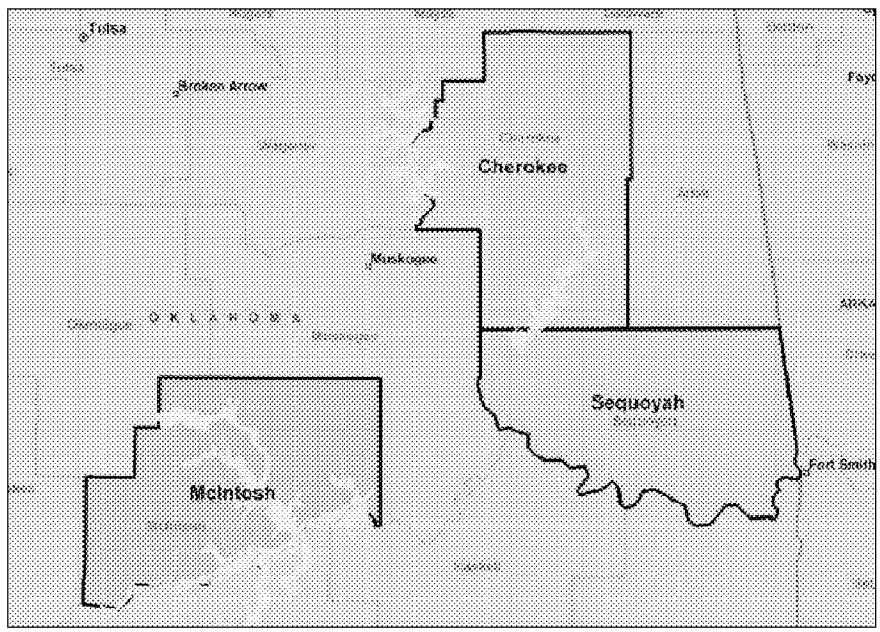
However, even if Eufaula Lake and Tenkiller Lake were not comparable lakes, i.e., there are characteristics that differentiate the two lakes, we would expect that as the alleged phosphorus problem worsened over time, the relative effect on home prices would be negative. For example, due to its size and location, Eufaula Lake may be windier and have less water clarity than Tenkiller Lake, which is "one of a handful of clear water lakes in Oklahoma," where as Lake Eufaula is not.[9] Based on this information, we might expect that a home located on Tenkiller Lake would have an increased value based solely on water clarity. However, if there was an increasing water quality problem over time at Tenkiller Lake that was not present at Eufaula Lake, then we would expect to find homes at Tenkiller Lake to appreciate at a slower rate than homes on Eufaula Lake, or possibly decline. In other words, as the (theoretical) water quality problem at Tenkiller Lake worsened, the effect on home price would become relatively negative.

[9] Lake Tenkiller, Oklahoma," www.outdoors.ok.come/Oklahoma/Tenkiller, Accessed March 26, 2009 and "Lake Eufaula, Oklahoma," www.outdoors.ok.come/Oklahoma/Eufaula, Accessed March 26, 2009.

3.1 Data

In order to test whether alleged pollution in Tenkiller Lake has had a negative effect on home price value, we collected sales transaction data on single family homes in neighborhoods located within a mile of Northwest Lake Eufaula (McIntosh County, OK) and Tenkiller Lake (Cherokee County and Sequoyah County, OK) from CountyAssessor.info.[10] Below is a map of the counties for which data were collected. Because there are areas of Eufaula Lake that have been identified as "impaired" by the Oklahoma Department of Environmental Quality[11] we only collected data for homes sold in McIntosh County, the county containing the northwest area of the lake.[12]

Figure 3.1: Counties for Which Data Was Collected



[10] County Assessor Offices represent that the data represents sales transaction data (where there is a buyer and seller) and not assessment data. However, there were 73 instances where a single property had more than one transaction in a single year; these transactions were excluded.

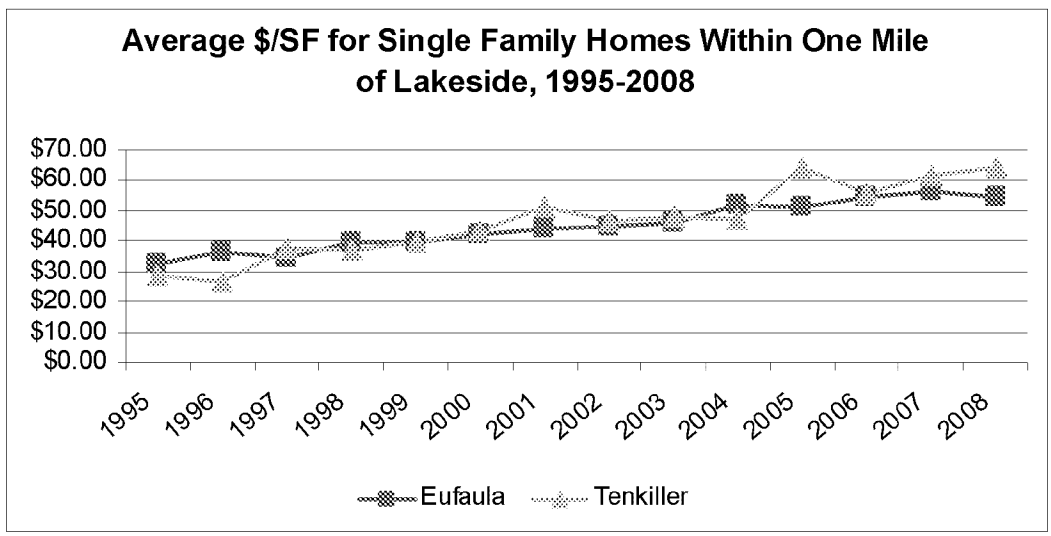
[11] Eufaula Lake, Canadian River Arm, and Longtown Creek Arm are identified as being impaired due to "Oxygen, Dissolved." In addition, the Canadian River Arm is identified as being impaired due to Turbidity and Color. The source of all impairments is "Unknown" (140). See Appendix C to Oklahoma Department of Environmental Quality. 2008. "Water Quality in Oklahoma: 2008 Integrated Report."

[12] Note that there are some areas in the southern area of McIntosh County that Plaintiffs may allege to be "Substantially Affected by Poultry Operations." Therefore, we have conducted a sensitivity analysis including and excluding the subdivisions located in this area. The results presented here are consistent whether or not properties located in southern McIntosh County are included.

Sales Price per Square Foot, 1995-2008

Plotting the price per square foot for Tenkiller Lake and Lake Eufaula shows that sales price per square foot was very similar at the two lakes between 1995 and 2008.[13] In addition, price per square foot increased at approximately the same rate during that time period. As noted above, the previous literature (Gibbs, et al. 2002; Michael, Boyle, and Bouchard 2000) has shown that decreases in water quality, measured through changes in the clarity of the water, will lead to reduced property values, especially for lakefront property. Based on our comparison, we do not find any appreciable difference between Tenkiller Lake and Lake Eufaula. The chart below reports the data within one mile of each of the respective lakes for 1995 to 2006.[14]

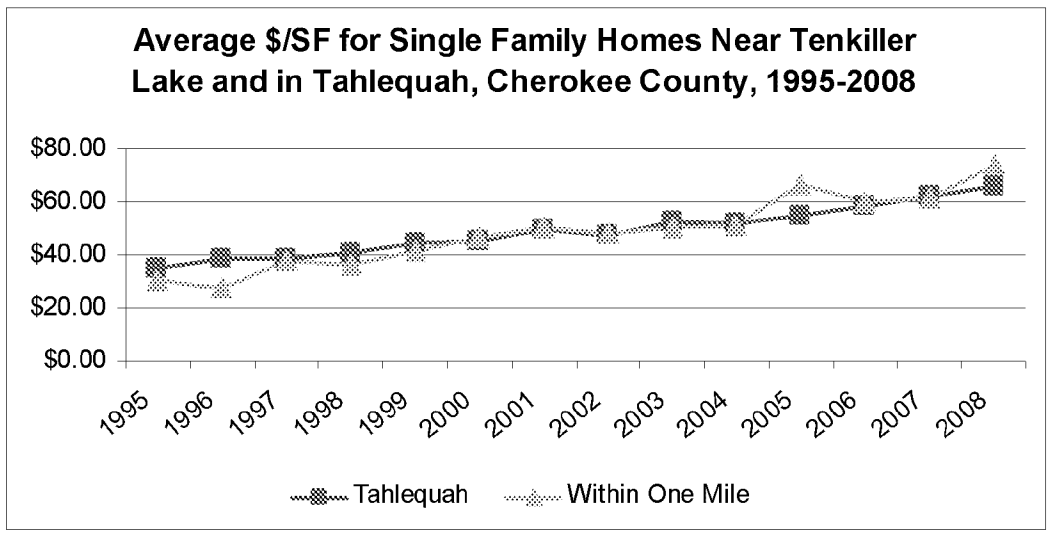
Figure 3.2: Average $/SF for Single Family Homes within One Mile of Lake, 1995 - 2008



Similarly, comparing the average price per square foot for homes within a mile of the lake to homes in nearby Tahlequah shows that price per square foot was similar in magnitude and change during 1995-2008.

[13] Although there was transaction dating back to 1987 for Lake Eufaula, there are less than 20 transactions per year for Tenkiller Lake until 1995.
[14] Results are very similar for homes within a half mile of each lake.

Figure 3.3: Average Price per Square Foot for Homes Near Tenkiller Lake and Homes in Tahlequah, Cherokee County, 1995 - 2008



3.2 Comparative Value of Tenkiller versus Eufaula Lake

There is wide variation of attributes of the single family home. Hence, it is necessary that we control for these various attributes that are expected to affect the transaction price of single family homes located near each of the two lakes. These attributes include square feet, the number of bedrooms and bathrooms, the condition of the building (Average, Excellent, Fair, Good, or Poor), age of the house at sale date,[15] air/ventilation type (forced air, heat pump, none, window unit, zoned F/A), and time. By using standard regression analysis, we found that the coefficient for Tenkiller Lake is positive and significant (at the 10% level) for homes within one mile of the lake, while the coefficient is positive, but not significant for homes located within a half mile of the lake. In other words, the effect on sales prices for properties within both one half and one mile on Tenkiller Lake for the past fourteen years is positive relative to comparable properties on Eufaula Lake and significantly positive for homes within one mile of the lake. Accordingly, there is no evidence, based on actual market transactions, that water quality has negatively impacted the valuation of single family homes on Tenkiller Lake.

Regression results are shown below. While the coefficient ("coef.") shows the effect (i.e., positive or negative) of the independent variable on the log of price, the p-value ("P>t") indicates whether the effect is statistically different from zero. Almost all of the independent variables are significant to the 10% level except for log_bedrooms, log_sale_age, and condition_3 ("fair" condition relative to "poor" condition) in the model for homes within one mile of the lake (Table 3.2). For homes within a half mile of the lake (Table 3.3), condition_3 becomes significant while lake_tenkiller is no longer significant.

Table 3.2: Regression Results for Homes Sold between 1995 and 2008 within One Mile of the Lake[16]

Number of obs = 1348						
F(15, 1332) = 105.83						
Prob > F = 0.0000						
R-squared = 0.5471						
Root MSE = .49416						
		Robust				
log_price	**Coef.**	**Std. Err.**	**t**	**P>t**	**[95% Conf. Interval]**	
log_sf	0.712	0.052	13.6	0.000	0.609	0.815
log_bedrooms	-0.022	0.066	-0.34	0.737	-0.151	0.107
log_bathrooms	0.291	0.061	4.77	0.000	0.171	0.411
log_sale_age	-0.023	0.023	-1.02	0.307	-0.068	0.021
lakeside_half	0.295	0.051	5.8	0.000	0.195	0.394
ventillation_1	0.451	0.072	6.26	0.000	0.310	0.593
ventillation_2	0.516	0.133	3.87	0.000	0.255	0.778
ventillation_4	0.197	0.071	2.78	0.006	0.058	0.336
ventillation_5	0.444	0.141	3.15	0.002	0.167	0.721
condition_1	0.644	0.159	4.06	0.000	0.333	0.956
condition_2	0.681	0.215	3.17	0.002	0.259	1.103
condition_3	0.261	0.164	1.59	0.111	-0.060	0.583
condition_4	0.857	0.163	5.26	0.000	0.537	1.176
Year	0.039	0.003	11.18	0.000	0.032	0.046
lake_tenkiller	0.062	0.036	1.74	0.082	-0.008	0.132
_cons	-73.761	7.013	-10.52	0.000	-87.520	-60.003

[15] Note that if there was an "effective year built," this date was used instead of "year built" to account for any property improvements.

[16] Baseline for Ventilation Type is "None" and baseline for Condition is "Poor". Ventilation Type 1 represents "Forced Air", Ventilation Type 2 represents "Heat Pump", Ventilation Type 4 represents "Window Unit" and Ventilation Type 5 represents "Zoned F/A". Condition 1 represents "Average", Condition2 Represents "Excellent", Condition 3 represents "Fair", Condition 4 represents "Good."

Table 3.3: Regression Results for Homes Sold between 1995 and 2008 within A Half Mile of the Lake

Number of obs = 1238						
F(14, 1223) = 92.06						
Prob > F = 0.0000						
R-squared = 0.5169						
Root MSE = .496						
		Robust				
log_price	**Coef.**	**Std. Err.**	**t**	**P>t**	**[95% Conf. Interval]**	
log_sf	0.721	0.054	13.45	0.000	0.616	0.826
log_bedrooms	0.000	0.070	0	0.996	-0.137	0.137
log_bathrooms	0.295	0.063	4.68	0.000	0.172	0.419
log_sale_age	-0.016	0.025	-0.66	0.509	-0.065	0.032
ventillation_1	0.455	0.079	5.76	0.000	0.300	0.610
ventillation_2	0.525	0.135	3.88	0.000	0.259	0.791
ventillation_4	0.199	0.078	2.55	0.011	0.046	0.352
ventillation_5	0.426	0.148	2.88	0.004	0.135	0.716
condition_1	0.679	0.164	4.13	0.000	0.356	1.001
condition_2	0.723	0.219	3.3	0.001	0.293	1.154
condition_3	0.292	0.171	1.71	0.088	-0.043	0.627
condition_4	0.891	0.169	5.29	0.000	0.560	1.221
Year	0.039	0.004	10.58	0.000	0.032	0.047
lake_tenkiller	0.053	0.036	1.47	0.142	-0.018	0.123
_cons	-74.281	7.475	-9.94	0.000	-88.945	-59.616

We also tested whether there was a significant change in the effect of a home being located on Tenkiller Lake over time. As previously stated, if the water quality at Tenkiller Lake was declining over time, we would expect the effect of being located on Tenkiller Lake to become negative over time. To test whether or not this was true, we introduced an "interactive term" based on a sale taking place near Tenkiller Lake and the year of the sale. The coefficient on this interactive term (tenkill_yr) measures the effect of being on Tenkiller Lake for the given year. While the interactive term is always positive, it has a *significantly* positive effect on sales price in 2000, 2001, and 2005 for homes within a mile of the lake (Table 3.4). Results are similar for homes within a half mile of the lake (Table 3.5). Thus, we find the evidence shows that sales prices on Tenkiller Lake compared to Lake Eufaula have not varied during the fourteen year time period, contradicting allegations made by plaintiffs that reduced water quality in Tenkiller Lake has resulted in a less desirable place in which to live and recreate.

Moreover, this would further indicate that there is no observed decline in value due to allegedly lower water quality on Tenkiller Lake.

Table 3.4: Regression Results for Homes Sold between 1995 and 2008 within One Mile of the Lake[17]

Number of obs = 1348						
F(28, 1319) = 59.05						
Prob > F = 0.0000						
R-squared = 0.5545						
Root MSE = .49255						
		Robust				
log_price	**Coef.**	**Std. Err.**	**t**	**P>t**	**[95% Conf.**	**Interval]**
log_sf	0.720	0.053	13.69	0.000	0.617	0.823
log_bedrooms	-0.018	0.066	-0.28	0.782	-0.148	0.111
log_bathrooms	0.284	0.062	4.61	0.000	0.163	0.404
log_sale_age	-0.027	0.022	-1.22	0.222	-0.071	0.017
lakeside_half	0.293	0.050	5.84	0.000	0.195	0.392
tenkill_year95	0.092	0.214	0.43	0.666	-0.327	0.512
tenkill_year96	0.144	0.157	0.92	0.358	-0.163	0.452
tenkill_year97	0.068	0.178	0.38	0.703	-0.282	0.418
tenkill_year98	(dropped)					
tenkill_year99	0.257	0.172	1.49	0.136	-0.081	0.596
tenkill_year00	0.373	0.159	2.35	0.019	0.061	0.685
tenkill_year01	0.406	0.174	2.34	0.019	0.066	0.747
tenkill_year02	0.252	0.169	1.49	0.138	-0.081	0.584
tenkill_year03	0.187	0.159	1.18	0.239	-0.124	0.499
tenkill_year04	0.112	0.204	0.55	0.582	-0.288	0.513
tenkill_year05	0.434	0.167	2.6	0.009	0.106	0.762
tenkill_year06	0.017	0.179	0.1	0.923	-0.335	0.369
tenkill_year07	0.052	0.191	0.27	0.785	-0.323	0.428
tenkill_year08	0.200	0.187	1.07	0.285	-0.167	0.567
ventillation_1	0.444	0.073	6.08	0.000	0.301	0.588
ventillation_2	0.525	0.142	3.69	0.000	0.246	0.803
ventillation_4	0.186	0.071	2.61	0.009	0.046	0.325
ventillation_5	0.445	0.142	3.12	0.002	0.165	0.724
condition_1	0.658	0.161	4.09	0.000	0.343	0.974
condition_2	0.684	0.216	3.17	0.002	0.260	1.107

[17] Baseline for Ventilation Type is "None" and baseline for Condition is "Poor". Ventilation Type 1 represents "Forced Air", Ventilation Type 2 represents "Heat Pump", Ventilation Type 4 represents "Window Unit" and Ventilation Type 5 represents "Zoned F/A". Condition 1 represents "Average", Condition2 Represents "Excellent", Condition 3 represents "Fair", Condition 4 represents "Good".

Number of obs = 1348						
F(28, 1319) = 59.05						
Prob > F = 0.0000						
R-squared = 0.5545						
Root MSE = .49255						
		Robust				
log_price	**Coef.**	**Std. Err.**	**t**	**P>t**	**[95% Conf. Interval]**	
condition_3	0.275	0.166	1.66	0.097	-0.050	0.601
condition_4	0.866	0.165	5.25	0.000	0.543	1.190
Year	0.039	0.004	10.18	0.000	0.031	0.046
lake_tenkiller	-0.126	0.136	-0.93	0.353	-0.394	0.141
_cons	-73.061	7.640	-9.56	0.000	-88.048	-58.074

Table 3.5: Regression Results for Homes Sold between 1995 and 2008 within A Half Mile of the Lake

Number of obs = 1238						
F(27, 1210) = 49.52						
Prob > F = 0.0000						
R-squared = 0.5247						
Root MSE = .49491						
		Robust				
log_price	**Coef.**	**Std. Err.**	**t**	**P>t**	**[95% Conf. Interval]**	
log_sf	0.729	0.054	13.51	0.000	0.623	0.835
log_bedrooms	0.004	0.070	0.05	0.958	-0.134	0.141
log_bathrooms	0.288	0.064	4.53	0.000	0.163	0.413
log_sale_age	-0.021	0.024	-0.87	0.386	-0.069	0.027
tenkill_year95	0.075	0.218	0.34	0.731	-0.354	0.504
tenkill_year96	0.125	0.164	0.76	0.445	-0.196	0.447
tenkill_year97	0.049	0.186	0.26	0.793	-0.316	0.413
tenkill_year98	(dropped)					
tenkill_year99	0.236	0.178	1.32	0.186	-0.114	0.586
tenkill_year00	0.347	0.166	2.09	0.037	0.022	0.673
tenkill_year01	0.370	0.183	2.02	0.044	0.011	0.729
tenkill_year02	0.223	0.176	1.27	0.205	-0.122	0.568
tenkill_year03	0.123	0.164	0.75	0.454	-0.199	0.445
tenkill_year04	0.083	0.210	0.39	0.693	-0.328	0.494
tenkill_year05	0.373	0.173	2.15	0.032	0.032	0.713
tenkill_year06	-0.016	0.186	-0.09	0.931	-0.381	0.348
tenkill_year07	-0.018	0.198	-0.09	0.926	-0.408	0.371
tenkill_year08	0.163	0.198	0.82	0.410	-0.226	0.552
Ventilation_1	0.449	0.080	5.61	0.000	0.292	0.606

Number of obs = 1238						
F(27, 1210) = 49.52						
Prob > F = 0.0000						
R-squared = 0.5247						
Root MSE = .49491						
		Robust				
log_price	**Coef.**	**Std. Err.**	**t**	**P>t**	**[95% Conf. Interval]**	
ventillation_2	0.533	0.144	3.69	0.000	0.250	0.817
ventillation_4	0.189	0.079	2.41	0.016	0.035	0.343
ventillation_5	0.432	0.150	2.89	0.004	0.139	0.726
condition_1	0.697	0.166	4.19	0.000	0.370	1.023
condition_2	0.725	0.221	3.29	0.001	0.293	1.158
condition_3	0.309	0.172	1.79	0.074	-0.029	0.647
condition_4	0.905	0.171	5.3	0.000	0.570	1.240
Year	0.040	0.004	9.75	0.000	0.032	0.048
lake_tenkiller	-0.101	0.144	-0.7	0.486	-0.383	0.182
_cons	-75.150	8.209	-9.15	0.000	-91.255	-59.045

3.3 References

Bureau of Labor Statistics. 1980-2009. "Consumer Price Index – All Urban Consumers, CUUR0300SA0, South urban, All items, Base Period 1982-84=100."

County Records, Inc., PvPlus Database of Assessor's records acquired from the County Offices of Cherokee County, Sequoyah County, and McIntosh County in Oklahoma. Accessed March 25-26, 2009 from http://countyassessor.info.

Entire Lakes Report Executive Summary, "Eufaula Lake," Oklahoma's Beneficial Use Monitoring Program – Lakes Sampling, 2006-2007 Draft Report, pp. 184 – 194.

Gibbs, Julie P., John M. Halstead, Kevin J. Boyle, and Ju-Chin Huang. 2002. "An Hedonic Analysis of the Effects of Lake Water Clarity on New Hampshire Lakefront Properties." *Agricultural and Resource Economics Review* 31(1): 39-46.

"Lake Eufaula, Oklahoma," www.outdoors.ok.come/Oklahoma/Eufaula. Accessed March 26, 2009.

"Lake Tenkiller, Oklahoma." www.outdoors.ok.come/Oklahoma/Tenkiller. Accessed March 26, 2009.

Michael, Holly J., Kevin J. Boyle, and Roy Bouchard. 2000. "Does the Measurement of Environmental Quality Affect Implicit Prices Estimated from Hedonic Models?" *Land Economics* 76(2): 283-298.

Oklahoma Department of Environmental Quality. 2008. "Water Quality in Oklahoma: 2008 Integrated Report." Appendix C.

Poor, P. Joan, Keri L. Pessagno, and Robert W. Paul. 2007. "Exploring the hedonic value of ambient water quality: A local watershed-based study." Ecological Economics 60: 797 - 806.

4. CRITIQUE OF CONTINGENT VALUATION SURVEY

This section describes the numerous biases in the Stratus CV survey questionnaire and key survey findings. As shown above, only after finding the unsatisfactory results (from their perspective) using the methods based on unbiased estimates of actual behaviors, did the plaintiffs' experts turn to the CV methodology. This methodology has been shown to be subject to substantial hypothetical biases, especially for passive users or nonusers, stemming from flaws in the survey questionnaire, as well as the survey administration. The Stratus CV survey describes a hypothetical referendum for a restoration project that would restore water clarity and ecosystem services to levels purported to have existed in the 1960s. Two important questions that we consider about this hypothetical referendum are whether the description is consistent with the appropriate conceptual economic underpinnings and whether respondents processed the description in the way that the analysts intended. Of course, that intention should be such that people's responses are elicited in a way that minimizes the potential for bias. The Stratus questionnaire and survey embodies numerous critical flaws in the description of the hypothetical commodity that render the survey responses invalid and the results unreliable for use in a damage assessment. Specifically, our analyses demonstrate that:

- The CV survey questionnaire contains biased and misleading information.
- CV survey respondents are valuing a commodity other than a faster recovery of the algae conditions for the Illinois River and Tenkiller Lake.
- The CV results contain substantial hypothetical bias.
- The CV results are an artifact of the hypothetical bid structure.
- The CV survey results suffer from nonresponse bias.

Each of these flaws, among others, is discussed below.

4.1 The CV survey questionnaire contains biased, misleading, and factually incorrect information.

Maintaining neutrality in a questionnaire is of critical importance in any survey, but especially in surveys used in litigation. The survey literature contains many

examples of the effects of non-neutral wording in biasing results (Rea and Parker 2005; Rossi, Wright, and Anderson 1983). The NOAA Panel specifically addresses neutrality by emphasizing the importance of a conservative design (Arrow, et al. 1993). While no set protocol exists for determining what constitutes neutral and conservative language, researchers should, at the very least, present balanced and unbiased information when describing the environmental issues.

The Stratus CV survey purports to provide respondents with an impartial description of the background regarding water quality conditions in the Illinois River System and Tenkiller Lake. In fact, the Stratus CV questionnaire is anything but balanced and unbiased. Examples of the bias are littered throughout the questionnaire. Among the most egregious examples of bias in the Stratus survey are the representations of the safety and efficacy of the proposed alum restoration project. Specifically, the survey questionnaire states that:

> Alum is used to keep pickles crisp, and you can buy alum powder in the grocery store for many uses, including cooking and making "play dough" for children.
>
> If alum is put on land, it attaches to phosphorous in the soil to form harmless particles. When these particles wash into rivers and lakes, the particles sink to the bottom and do not help algae grow.
>
> For more than 35 years, alum has been used successfully and safely to remove phosphorous and reduce algae in many states, such as Colorado, Texas, Missouri, South Dakota, Florida, Wisconsin, and Washington…..Experiences in those states have convinced scientists that alum does not harm fish or other things living in water, and that alum treatments here in Oklahoma could safely return the river and Lake to what they were like in around 1960.

Thus, the alum picture painted in the survey (reinforced by the grocery store photograph that shows alum powder in a small spice container next to other spices used by home cooks) is that the alum restoration program would be a safe and effective way to reduce algae in the Illinois River System and Tenkiller Lake.

The safety of alum for fish and other biota is a subject of considerable debate in the scientific community. Specifically, Connolly, Sullivan, and Coale (2009) cite

numerous references as to the potential risks to fish and other biota from the use of alum in a restoration project. Chief among these problems for fish and other biota are:

- Possible morphological deformities in benthic communities
- Diminished survival of some spring spawning fish and bottom-dwelling amphibians
- Possible chronic effects on fish.

Clearly, the Stratus questionnaire provides no mention of such potential risks to fish or other biota (p.18).

Connolly, Sullivan, and Coale (2009) further note the risks to forage grasses from the application of alum to pasture lands. They indicate that the science of alum restoration for forage grasses is anything but well-developed and that such a large-scale program as proposed in the questionnaire would raise significant technical issues, especially ones related to substantial changes in the acidity of the soils that would require application of other minerals on a large scale to offset the alum impacts. Moreover, Connolly, Sullivan, Coale (2009) indicate that determining the rate of application for alum and the other minerals would have to be done on a field-by-field basis because of the diversity of acidity levels in the soil. Finally, the Stratus questionnaire fails to mention that the alum application would take place on private lands, which would raise significant implementation problems for the hypothetical program. Alum restoration on land would pose substantial risks to forage grasses, which are critical to the economic well-being of the farmers who raise cattle in the Illinois River watershed. None of these risks, or the potential economic trade-offs that may be associated with an alum restoration program, is described in the survey questionnaire.

Connolly, Sullivan, and Coale (2009) indicate that alum restoration projects in other locations have been the subject of considerable controversy among various interest groups, especially nearby residents. They cite case studies in which proposed restoration projects were either delayed for several years, or modified because of public concerns about the safety to fish and shellfish. The Stratus CV survey designers

presented none of this information about the reaction in other communities to alum restoration projects in the survey questionnaire for the Illinois River System and Tenkiller Lake.

Similarly, the questionnaire fails to reflect the unknown efficacy of the proposed alum treatment. Instead, the implementation of the alum restoration program is described in almost trifling simplicity:

> Alum could be spread on land from trucks.
>
> Alum could be spread on the lake from boats.
>
> Alum could be sprayed in river water flowing into Oklahoma from Arkansas.

This restoration program is largely a figment of the survey designers' imagination, not the depiction of a realistic restoration option. Perhaps, the most telling refutation of the alum restoration program comes from the plaintiffs' own restoration consultant, Mr. King. Specifically, in his report, Mr. King states (King 2008, p. 19):

> However, in a reservoir, such as Lake Tenkiller, high dosages and repeated applications may be needed to be potentially effective in sequestering sediment P. With higher dosages, there is the potential for localized depression of pH with an associated potential increase in aluminum toxicity to aquatic life.
>
> Alum treatment of Lake Tenkiller could potentially reduce the internal loading of P from lake sediments. Using alum typically increases the water clarity. Alum can be toxic to aquatic life at low pH (Cooke et al., 2005). Alum applications are generally effective in lakes from 5 to 15 years (Welch and Cooke, 1999). However, the duration of alum treatment effectiveness in a reservoir such as Lake Tenkiller will not be as long as a lake and will be further reduced proportional to the additional P inputs from the Illinois River, Caney Creek and the Baron Fork. Therefore, the applicability of P inactivation with alum cannot be adequately evaluated until the final remedial measures for the watershed and riverine response regions have been identified in sufficient detail to determine future P and nutrient loadings to Lake Tenkiller.

When asked about alum restoration in his deposition, he indicated that he had rejected the option because it was not technically feasible. Specifically, Mr. King states (King 2009, pp. 287-288):

> Q. On Page 19, one of these potential treatments you discussed is P inactivation with alum, aluminum sulfate; correct?
>
> A. Yes, sir.
>
> Q. This specific potential remedy or remedial step is one that you are not recommending to be implemented at this time; is that correct?
>
> A. I categorized it as requires additional investigation and assessment.
>
> Q. And does that mean that you cannot recommend it at this time based upon the current data in hand?
>
> A. Yes.
>
> Q. To your knowledge, has anyone done a technical evaluation of the feasibility of treating Tenkiller Reservoir with alum?
>
> A. No, no, not that I can think of.

In addition, the Stratus survey contained "scientific" information about the effects of algae on fish in the Illinois River and Lake Tenkiller. This scientific information is presented in such a way as to convey that there is no scientific debate about the accuracy of the information.[18] The key statements include the following:

- Fewer small mouth bass, other fish and small plants in both the IR and Tenkiller Lake
- Large areas of Tenkiller Lake small mouth bass and other types of fish people catch grow slower and there are fewer of them
- Large areas of the bottom of Tenkiller Lake, there are lot fewer insects and small animals than are in the lakes with less algae
- Large mouth bass have increased in numbers and growing more quickly.

[18] Of course, the questionnaire designers note in the survey that scientists agree that the effects of algae were the result of human activities (p. A-13.) Such a statement is so broad as to be meaningless. Nevertheless, it conveys the impression that scientists agree with all the other information that is presented in the survey, which is inaccurate.

However, Connolly (2009) offers a very different picture as to the impacts of phosphorous on fish populations in the Illinois River and Tenkiller Lake. For example, he concludes:

- The fish community within the Illinois River Watershed is not highly degraded due to water quality impacts. Lower diversity is more of a function in stream-size than reduced water quality.
- Lower diversity is more affected by poor stream habitat than water quality.
- The sample protocols may underestimate the diversity of fish in the Watershed.
- One would expect the bass fishery in Tenkiller Lake to be dominated by largemouth bass, followed by spotted bass, with small mouth bass a minor component due to the habitat requirements of the latter species.

In his deposition, Dr. Cooke, one of the plaintiffs' biological consultants notes that the construction of a dam had a significant impact on small mouth bass in the Illinois River. Specifically, Dr. Cooke states (Cooke 2008, pp. 557-558):

> Q. Now, you say in your report that smallmouth bass were abundant in the Illinois River – excuse me, David -- prior to the formation of the lake?
>
> A. Yes.
>
> Q. And when the reservoir was formed, would you and the dam was closed, the lake began to fill, would you agree with me that that created a very different habitat, fish habitat than the flowing Illinois River watershed?
>
> A. I would agree with that.

The CV survey fails to mention any potential impacts from the construction of the dam on the small mouth bass in the Illinois River and Tenkiller Lake, nor does it mention anything about differences in habitat requirements of the various bass species.

The failure to acknowledge any uncertainty among scientists about the potential injuries or the safety and efficacy of alum in the survey questionnaire is a serious flaw. By not reflecting the scientific uncertainty associated with the injury and the restoration program in the survey questionnaire, the survey adds another dimension for biasing the

survey results to generate a higher damage estimate. The existing literature on uncertainty clearly demonstrates that including such information would have substantially altered the responses. Specifically, it has long been known from the psychological literature that people have a very difficult time answering questions where uncertainty is present. In particular, the literature shows that people's preferences are often poorly formed, are very sensitive to the way questions are framed, and that people are unable to process probabilistic information (Tversky and Kahneman 1981; Slovic, Fischoff, and Lichtenstein 1982). One research finding that is particularly pertinent to the Stratus CV questionnaire is the so called "certainty effect" (Weinstein and Quinn 1983; Tversky and Kahneman 1981). People respond to questions quite differently when one of the options presented involves a certain outcome. Thus, the Stratus questionnaire presents respondents with a biased, inadequate basis for evaluating the hypothetical restoration program, rendering the results invalid.

Another facet of bias in the CV questionnaire involves the discussion of the poultry industry as the primary source of the algae growth. Specifically, the questionnaire tells respondents that "60 percent of the phosphorous in the IR and TenKiller Lake is from chickens and turkeys." This statement, the accuracy of which is attributed to Dr. Engel's various reports, is of critical importance to the survey designers. Without it, they have no way to associate the phosphorous loads to the Illinois River and Tenkiller Lake with the application of poultry litter.[19] As Connolly, Sullivan, and Coale (2009) state, Dr. Engel's methodology that produces the 60 percent estimate is without scientific foundation. In addition, Dr. Bierman (2009) concludes that Engel's approach is an inappropriate tool for predicting watershed nonpoint source phosphorous loads. Dr. Bierman further concludes that Engel's approach is inconsistent with accepted practices in the scientific community and that it contains numerous and substantial errors. Survey respondents are provided none of this

[19] The survey questionnaire fails to explicitly tell people what will happen to the other forty percent of future phosphorous loads to the Illinois River. Survey respondents likely derived the impression that the combination of alum treatments and the ban on poultry applications would remove both the past and future phosphorous loads from all sources. Of course, if respondents believed such an outcome would occur, their votes would be based on a perception that exceeds the scope of the injury alleged by the plaintiffs in this case.

information about the questions concerning the scientific validity of Dr. Engel's 60 percent estimate.

Moreover, the questionnaire contains a specific and detailed focus on poultry litter as the cause of the algae. The questionnaire emphasizes the role of the poultry industry by enumerating that 140 million chickens and turkeys are raised each year within the watershed and that these birds produce more than 300,000 tons of litter annually. Other than mentioning, in passing, that the 40 percent of phosphorus attributable to other sources includes sewage treatment and store-bought fertilizer applications and the cattle industry, there is no comparable specificity for these other sources. That is, the questionnaire is silent on the number of individual septic fields within the watershed, the number of households served by sewage treatment facilities within the watershed, and the number of acres of lawns and golf courses to which store-bought fertilizer is applied (among other potential sources of phosphorus). The lack of specificity about the other sources of phosphorus results in an unbalanced and biased questionnaire.

The restoration recovery periods are a critical component of the hypothetical scenarios in the Stratus survey questionnaire. The survey describes the natural recovery for the river and the lake once the ban on poultry litter application was imposed. However, Connolly, Sullivan, and Coale (2009) conclude:

> As a result, the statements by Stratus in their Survey that the river and lake would recover to 1960's conditions in about 60 and 50 years, respectively, once poultry litter application was stopped, can not be supported. The models developed by the Plaintiffs can not provide an accurate measurement of this "time to recovery" as they are currently developed and applied. (p.13)

Moreover, Connolly, Sullivan, and Coale (2009) further conclude that there is no scientific basis for the survey's contention that the alum restoration program would speed the return of water quality to its purported historical levels. Specifically, they state:

> However, no scientific basis is given in Chapman et al. (2009) for this 40-year acceleration. The one citation given in the Chapman et al. (2009) report for alum treatment in the watershed (Cooke et al. 2005) actually states that alum treatment of reservoirs is uncommon and somewhat discourages direct application of alum in flowing rivers (see Section 3 of this report for further discussion). Nowhere in Cooke et al. (2005) is information provided that would allow one to quantify the acceleration of recovery using alum (p.13).

In terms of balance, omitting relevant counter-arguments from the text can further bias the respondent. Schuman and Presser (1981) find "that the effects of adding counter-arguments are too pervasive and too large to allow the question forms…to be treated as interchangeable…." In other words, adding counter arguments provides such great changes in response outcomes that questions which provide counter arguments cannot be treated as identical to questions which do not provide counter arguments. The counter argument gives the respondent who has not previously considered an issue a plausible reason for choosing the other side of the issue. Schuman and Presser state: "The counter argument thus provides a genuine degree of cognitive persuasion, and is not merely a matter of social pressure." Specific counterarguments for the restoration program might be that the program has not been fully evaluated by scientists and the potential economic tradeoffs in the form of higher costs to farmers who grow hay as well as cattle ranchers.

Respondents' open-ended comments indicate that this questionnaire was not sufficiently balanced in terms of counter-arguments. Near the end of the survey, the questionnaire asks respondents whether they felt pushed to vote in a certain direction. Despite the almost hour long in-person interview dosing respondents with information about water quality impacts from the poultry industry, almost 9 percent of the respondents to the base questionnaire admitted that they felt pushed to vote for the alum treatments.[20] When asked why they felt this way, they responded:

- "Because it totally disregarded other things in the land and just spoke of alum and phosphorus"

[20] This percent likely understates the percentage of respondents who felt pushed because of their unwillingness to express opinions that the interviewers might have viewed as being critical of the survey. This is another indication of the tendency that respondents have to want to please survey interviewers discussed above.

- "Because it does not discuss the socio-economic ramifications as in the poultry farmers, the communities supported by jobs in the poultry farms, monies lost by businesses like corn seed, doesn't discuss the higher cost of food such as poultry"
- "Excess information about the treatment"
- "It seems one-sided. The State wants to do it, so it's pushing for the alum treatments."
- "It seemed to only offer evidence to positive effect, but it didn't seem to offer any side effects to the contrary."
- "It did not provide enough contradictory information regarding the alum treatments."
- "Just showed one side."
- "It seemed to only offer evidence to positive effect but didn't seem to offer any side effects to the contrary."
- "Gave a more positive picture of the alum treatments than not."
- "I think I heard only one side of the story."
- "This was a state infomercial."
- "The statements did seem slanted towards the alum treatments. If I had not heard, I probably would have voted against them."
- "Most of the information was positive for the alum treatments. I would like to hear about other states that have used and any other side effects from it."
- "Because the opinions of the opposite parties involved were not included."
- "I didn't want to vote for something that would hurt farmers and thought it emphasized poultry litter too much, not 60%. I thought that the sewage and chemical fertilizer might affect the river more."
- "The pictures are taken to specially convince me about the algae. The picture cards e, f, and g are taken to make me vote for them."
- "That's why they are spending all this money to send you all here. So we will vote for a tax increase."

Clearly, based on the responses above as well as the other arguments described earlier, the Stratus questionnaire is seriously deficient in presenting counter arguments.

The photos used to depict the increase in algae are also relevant to the discussion of neutrality and conservative design (Mathews, Freeman, and Desvousges

2007; Arrow, et al. 1993; Mitchell and Carson 1989). Recall the respondent comment above that the photos were "specially taken to convince me about the algae." Because "a picture is worth a thousand words," photos are efficient survey tools. That efficiency is accompanied by the creation of an indelible image in the minds of the respondents. Although the Stratus team claims to use photos that show "relatively mild" algae growth, the differences are striking. They are so striking, in fact, that it is easy to forget that those conditions, where they exist in the river, are present only during a few months of the year and confined to limited areas. The interviewers only verbally mentioned these seasonal and spatial differences, making them easier to forget than the images presented in the photos. Moreover, the questionnaire fails to provide any details on how limited the areas might be that are represented in the photographs. Card N, which provides some reasons why the respondent might choose to vote against the alum treatments, is also silent on both the limited seasonal and spatial algae impacts.[21] A more neutral approach, to provide balance against the photos' lasting impressions, would have included both the seasonal and spatial limits on the algae in the photos and would have reminded respondents of these limits just prior to voting as a reason to potentially vote against the program.

Another critically important but biased facet of the Stratus questionnaire is the statement that asks respondents to assume that the Court had decided to impose a ban on the application of poultry litter in the Illinois River watershed. Such a statement is likely to indicate that the Court had already sanctioned the ban, when in fact the Court decided not to impose the temporary injunction sought by the plaintiffs in this case. The likely effect of such a statement is to mislead people to think that the Court agreed that the application of litter was a serious problem. Otherwise, it would not have been stopped. Such a misleading statement imports significant bias making it more likely that respondents would vote for the hypothetical restoration program.

Notifying the respondents of the sponsor of the survey, such as the use of the introductory letter from the State of Oklahoma, may cause them to respond as they believe the sponsor would like them to answer. Presser, Blair, and Triplett (1992) find a

[21] Connolly, Sullivan and Coale (2009) express criticisms of the photos from a scientific water quality perspective as well.

significant change in response distribution when the sponsor is named. They hypothesize that this result reflects the conjunction of two factors. First, respondents perceive that the sponsor had taken a clear position on the subject in question. In addition, the issue was one on which it was likely the respondent had not already formed an opinion. This CV survey clearly exhibits both of the qualifications that Presser, Blair, and Triplett (1992) hypothesize to be important: (1) the State clearly has an opinion on this subject or they would not be sponsoring the survey, and (2) because the scenario is hypothetical, respondents could not have previously formed an opinion.

Results from the earlier telephone survey conducted by Stratus likely influenced the information content in the CV questionnaire. As described above in Section 2.2, Stratus conducted a telephone survey of Oklahoma residents in 2006 to assess the knowledge and use of the Illinois River System and Tenkiller Lake, to determine perceptions about water quality, and to identify any impacts from media coverage of the environmental issues within the watershed (Stratus 2007). Table 4.1 below provides the progression of the questions asked about respondents' impressions of the Illinois River System and Tenkiller Lake.

Table 4.1: Respondents' Impressions and Knowledge about Tenkiller Lake and the Illinois River from 2006 Telephone Survey

Survey Question	Percent of Respondents mentioning poultry litter
"What impression do you have about the Illinois River? Is there anything especially good or bad about the Illinois River?"	6%
"What impression do you have about Tenkiller Lake? Is there anything especially good or bad about the Lake?"	2%
"Have you heard of any issues or concerns relating to the Illinois River or Tenkiller Lake or are you unaware of any issues or concerns there?"	16%
"What about water quality in the Illinois River and Tenkiller Lake? Are you aware of any water quality problems there or have you not heard of any water quality problems?"	26%

As those results show, increasing the amount of prompting and information used in the question can alter the responses to the survey. Respondents tend to agree

and provide the interviewer with the information they are looking for, which may not necessarily reflect their true impressions or opinions. As more information is given to prompt the respondent to provide information about water quality problems, more respondents comply and recognize the issue in their responses. After the telephone survey results were shared with the Stratus team, one member commented: "If estimated damages are to be significant, people will have to be educated about the injuries. There is currently not a lot of knowledge of the injuries" (Morey 2006).

Rather than first ask the 2008 CV respondents the extent of their knowledge and impressions in a manner similar to the 2006 telephone survey, the 2008 CV questionnaire *first* described the environmental issue as viewed by the plaintiffs and *then* asked respondents whether they had heard about these issues. Almost one-third of the 2008 respondents indicated that they had heard about the algae. This higher response may be due in part to respondents not wanting to appear uninformed about issues in their state. It is possible that the increasing media coverage of the Illinois River watershed and the Attorney General's lawsuit has raised awareness. Nevertheless, Stratus chose to not ask the 2008 respondents their impressions prior to "educating" them. Not doing so is inconsistent with a conservative design required by the NOAA Blue Ribbon Panel. Not doing so makes it impossible to disentangle potential nonuse values that respondents may have held prior to taking the survey and the nonuse values that were created during the "education" process that occurred in the CV survey. Thus, the Stratus questionnaire has artificially inflated, and in some instances created, the concerns about water quality in the Illinois River System and Tenkiller Lake by dosing the respondents with new (and, in some cases, flawed and erroneous) information before eliciting their opinions.

4.2 Many survey respondents valued a different commodity than was intended by the survey designers, rendering the results invalid.

A critical requirement for a CV survey is to provide information to respondents about the commodity so that they understand and accept it and can give a meaningful answer to the valuation question. The Stratus CV survey results reveal that respondents did not understand or accept the information in the CV scenario and thus did not value the commodity they were being asked to value—the return of water clarity

and ecosystem services to levels that purportedly existed in the 1960s. As a result, the CV responses do not reflect the value for water clarity and ecosystem service improvements in the Illinois River and Tenkiller Lake and are, therefore, meaningless.

The problem of information provision and acceptance is a basic concern with hypothetical CV results. The NOAA Panel expressed concerns about information provision and acceptance in its 1993 report:

> If CV surveys are to elicit useful information about willingness to pay, respondents must understand exactly what it is they are being asked to value (or vote upon) and must accept the scenario in formulating their responses (Arrow, et al. 1993, p. 4605).

The NOAA Panel's main concern is that respondents sometimes do not value the commodity specified in the survey that researchers assume they are valuing either because of commodity misspecification or scenario rejection. While the distinction between lack of understanding of the commodity and scenario rejection is frequently difficult to make, both problems have the same consequence: respondents are not valuing the commodity that researchers assume they are valuing. As a result, responses to the valuation question are difficult to interpret.

The Stratus CV Survey included several questions following the vote question to assess the respondents' acceptance of the "facts" described by the interviewer. These questions included the respondents' beliefs or understanding about:

- Whether the alum treatments would be implemented without the court-ordered ban
- Whether phosphorus had caused the changes described (or whether the respondent did not believe that the described changes had actually occurred)
- Whether the natural recovery period of the lake or river was different from that stated by the interviewer
- Whether the tax amount paid by each household would be different from that described by the interviewer
- Whether the tax collected would be used to clean up other lakes and rivers in addition to the Illinois River and Tenkiller Lake.

These data reveal that more than 80 percent of the respondents who voted for the program in the base version of the survey rejected at least one element of the scenario. Nearly 55 percent of them rejected at least two elements, and more than 25 percent of them rejected at least 3 elements. Table 4.2 contains the details.

Table 4.2: Scenario Acceptance Data

	Base Version Respondents Who Voted For	
Category of rejected elements	**Number of Respondents**	**Percent of Respondents**
Different Natural Recovery times	399	62%
Other lakes and rivers would also be cleaned up	289	45%
Different Tax Amount	270	42%
Alum Treatments might be implemented without the court-ordered ban	130	20%
Phosphorus had not caused the described changes	36	6%
TOTAL	**524**	**81%**
Number Rejecting 2 Elements	183	28%
Number Rejecting 3 Elements	122	19%
Number Rejecting 4 or 5 Elements	54	8%

Two of the results merit further discussion. About 45 percent of the base version respondents who voted for the alum treatments believed that the extra tax money would be used for cleaning up other lakes and rivers. Thus, almost half of the respondents were valuing a much larger commodity when they cast their hypothetical votes.[22] This is not the first time that respondents have not followed the exhortations of the survey designers to only value a specific resource. For example, in the Clark Fork River Basin contingent valuation study, the survey designers went to considerable lengths to inform respondents that their answers would only apply to resources in that river basin. However, when they asked people whether they considered only the Clark

Fork River Basin in developing their answers, approximately 83 percent indicated that they were valuing something other than the Clark Fork River (Diamond and Hausman 1994). Because many survey responses and votes reflect more than the Illinois River System and Tenkiller Lake, the resulting willingness-to-pay (WTP) estimates are not valid.[23] Specifically, they do not "fit" or correspond to the alleged natural resources injuries for the Illinois River System and Tenkiller Lake.

Moreover, respondents formed different assumptions about future phosphorus loads to the river and lake. Although they were told to assume that the alum treatments would occur only if the court ordered a ban on poultry litter spreading, clearly some respondents did not assume that the ban would occur when they hypothetically voted. What these respondents assumed about future phosphorus loads from poultry litter is unknown. What these respondents assumed about how long the water in the Illinois River System and Tenkiller Lake would remain clear in the absence of a ban also is unknown. Moreover, these ambiguities leave one without any sense of what respondents assumed about the future phosphorus load from other sources. Although the questionnaire briefly mentions "other things being done" to reduce (but not eliminate) new phosphorus from other sources, at least some of the future phosphorus loading will continue from other sources. This further complicates the interpretation of what people think they are valuing in this survey. Specifically, some respondents may have thought through the logic and "facts" in the survey and assumed that future phosphorus loadings from other sources would continue, even with the ban and the alum treatments. For that group, the number of future years of water clarity they believed would be achieved likely varied widely because of the differences in beliefs about the number of years natural recovery would take. On the other hand, some respondents likely forgot that other future sources of phosphorus would continue because these sources received so little attention in the questionnaire. The likely conclusion is that these respondents assumed that the ban and alum treatment would result in water clarity into perpetuity. Thus, respondents were valuing different commodities when they hypothetically voted.

[22] Section 5 contains the results of an analysis that explores the impact on WTP of valuing this much larger commodity.

Further evidence can be found that some respondents simply rejected what the survey designers intended for them to believe. For example, about 5 percent of the respondents who voted "for" the alum treatments did so because they were motivated by human health concerns. In a natural resource damage assessment, values related to potential health effects are not included as part of the definition of compensable values. Natural resource damages apply only to the residual injuries to natural resources after remediation is completed. The regulations presume that health considerations are addressed in remediation decisions, not restoration (43 CFR Part 11).

The responses to open-ended comments provide some additional insight that respondents are concerned about the possible consequences on their health and the health of others. Specifically, respondents who voted for the program said that they did so because:

- "Health being a major thing."
- "Husband had an ear infection as a result of swimming at lake."
- "A couple of years ago there was an incident of a child dying."
- "It would help a lot more people not to get sick from swimming."

Additionally, the published literature on risk perceptions indicates that such perceptions are not easy to change, even if the questionnaire is silent on the issue of human health. The Schulze, et al. (1998) study on Denver air quality demonstrates how difficult it was to get people to focus on just the visibility aspects of air pollution. This study explains that respondents have a "mental model" of environmental effects. Specifically, respondents believe that improvements in air visibility must necessarily be accompanied by improvements in healthiness of the air and other public goods. As a result, respondents with such a mental model embed health values into their responses that were not part of the CV question. It is likely that some respondents used the same type of mental model when answering the valuation question in the Stratus CV study.

[23] According to the Stratus report, respondents who thought that the extra tax payments would be used for cleaning up more rivers and lakes were more likely to vote for the alum treatments. This result is evident in Table 6.26, which indicates the statistical significance of this variable in predicting the voting patterns.

That is, the respondents assumed that a change in the visual aspects of the water must correspond to improvements in human health. Accordingly, the CV estimates include yet another element that is not related to the compensable value of the natural resource services.

Additionally, interviewer evaluation data from the Stratus survey reveal that some respondents did not understand or did not take the interview and the vote seriously. Following the interview with the respondent, each interviewer answered questions related to his/her opinion of the respondent's understanding and cooperation during the interview process. The interviewer evaluation data reveal that the interviewers indicated comprehension or cooperation issues with 90 respondents who voted for the program in the base version of the survey. Table 4.3 contains the details. Despite these obvious flaws, Stratus did not eliminate these respondents from their damage calculations.

Table 4.3: Interviewer Evaluation Data

	Base Version Respondents Who Voted "For"	
Category	**Number of Respondents**	**Percent of Respondents**
Comprehension Problems	37	6%
Distracted	25	4%
Impatient	24	4%
Inattentive	21	3%
Did Not Take Interview Seriously	9	1%
TOTAL Respondents	**90**	**14%**

In summary, the differences in understanding and scenario acceptance will cause different respondents to value a different commodity. Because two respondents are valuing different scenarios, their answers will not be comparable. In light of differences in comprehension and acceptance, there is no way to know what assumptions the respondents are making as they answer the questions. Accordingly, there is no way to know what bias these unknown assumptions are introducing into the CV results. However, to the extent that many of these perception problems go beyond

the specified injuries in the survey questionnaire, the most likely impact is an upward bias in the Stratus damage estimates.

4.3 Hypothetical bias is a fatal flaw in the Stratus CV data.

CV results are not based on actual, observed behavior made by people in an economic market who face the consequences of their decisions. Instead, the results are based on verbal interviews asking unusual questions about potentially unfamiliar, hypothetical situations. If the respondents make different decisions in this hypothetical scenario than they would if faced with the actual situation, then the results will be unreliable and unusable for assessing damages.

The difference between stated intentions and actual behavior is a reflection of hypothetical bias. Researchers recognized hypothetical bias in CV studies nearly 30 years ago, defining it as the "potential error due to not confronting an individual with a real situation" (Rowe, d'Arge, and Brookshire 1980, p. 6).[24] In effect, the hypothetical nature of CV does not provide respondents with an incentive to reveal their true values because they do not have to bear the consequences of any answers they provide in a survey. Common sense suggests that people simply will not put forth the same effort in making a choice when the outcome does not affect them. It is basically the difference between window shopping and making actual purchases. Because the respondent does not actually pay the stated amount in a CV survey, there is no penalty for giving an answer different from the person's true preferences.[25]

[24] Hypothetical bias is not unique to CV studies, but can be found in other types of studies that rely on people's intentions rather than their actual behaviors. Kemp and Maxwell's (1993) review of marketing studies shows that stated intentions do not reflect actual purchases. Swait, Louviere, and Williams (1994) report that the hypothetical data can mispredict shipping company market shares by as much as 40 percent. Adamowicz, Louviere, and Williams (1994) demonstrate that anglers' values for a fishing trip based on hypothetical data was two to eight times higher than values based on actual fishing trips. Desvousges, MacNair, and Smith (2000) reveal that hypothetical stated preference techniques imply anglers state that they are willing to drive 158 miles to avoid fishing at a site with a fish consumption advisory, compared to actual trip data where they only drive 18 miles to avoid the same type of advisories.

[25] Harrison (2007) reveals that about 40 percent of the respondents who took the *Exxon Valdez* CV survey believed that the survey was part of the damage assessment. The Stratus CV survey did not ask the respondents a similar question. However, to the extent that the Stratus respondents guessed that the survey results have a role in the Attorney General's lawsuit, they may not have the proper motives to reveal their true WTP values.

The NOAA Panel concurred on the issue of hypothetical bias leading to overestimates of damages by stating:

> "The Panel is persuaded that hypothetical markets tend to overstate willingness-to-pay for private as well as public goods. The same bias must be expected to occur in contingent valuation studies" (Arrow, et al.1993, p. 4610).

In summary, hypothetical bias invalidates CV estimates of total value, which includes both use and nonuse values. Because respondents do not have the incentive to provide their true answers and do not bear the consequences of their responses, CV results are not economically sound when used in the manner proposed by Stratus and the plaintiffs.[26]

4.3.1 The Stratus CV survey results demonstrate hypothetical bias.

Results of the Stratus CV study indicate that the respondents exhibit response patterns consistent with hypothetical bias. Most strikingly, more than one-third of the base version respondents did not pay state income taxes in 2007 (Table 6.19). Of these respondents, more than 58 percent of them voted for the alum treatments. When a full refund is also factored in, 258 respondents of the 647 respondents who voted for the alum treatments did not pay state income taxes. Thus, for more than one-third of the "for" respondents, the commitment of dollars was entirely hypothetical. They voted for the alum treatments without any commitment or belief that they would have to pay the cost of the alum treatments. Certainly, hypothetical bias permeates these CV results.

The Stratus CV Survey permits the evaluation of several respondent opinions and beliefs, which would be at odds with voting for the alum treatments. These inconsistencies in logic are evidence of hypothetical bias. These questions included the respondents' beliefs and opinions that

- Decreasing water pollution in the State is not at all or only slightly important
- Decreasing state income taxes is very or extremely important

[26] Appendix A contains a detailed review of literature on hypothetical bias.

- A lot less State tax money should be spent on the environment
- The algae conditions in ***both*** the lake and river are not at all or only slightly serious
- ***Both*** the lake and river will naturally recover faster than what the interviewer indicated
- The alum treatments will not work well at all, or only slightly well
- The tax would be higher than that indicated by the interviewer
- University scientists cannot be believed at all, or only a little
- State government officials cannot be believed at all, or only a little
- They do not consider themselves to be environmentalists at all
- The extra tax amount would be extremely or very difficult for their households to pay

These data reveal that about 84 percent of the respondents who voted for the program in the base version of the survey demonstrate at least one of these inconsistencies in logic. More than 50 percent demonstrate at least two inconsistencies, and 23 percent of them demonstrate at least 3 inconsistencies. Table 4.4 contains the details.

Table 4.4: Inconsistencies in Respondents' Answers

	Base Version Respondents Who Voted For	
Category	**Number of Respondents**	**Percent of Respondents**
Decreasing State Income Taxes is Important	366	57%
State government officials cannot be believed	221	34%
The tax amount will be higher	173	27%
The lake and river will naturally recover faster	110	17%
The extra tax will be difficult to pay	77	12%
University scientists cannot be believed	48	7%
Not at all an environmentalist	42	6%
The algae conditions are not serious	28	4%

	Base Version Respondents Who Voted For	
Category	**Number of Respondents**	**Percent of Respondents**
The alum treatments will not work	28	4%
Decreasing Water Pollution is Not Important	20	3%
A Lot Less Should Be Spent on the Environment	4	1%
TOTAL	**544**	**84%**
Number inconsistent in 5 or more aspects	18	3%
Number inconsistent in 4 aspects	51	8%
Number inconsistent in 3 aspects	82	13%
Number inconsistent in 2 aspects	177	27%

To the extent that these statements represent the true beliefs of the respondents, it is illogical that they would, in reality, agree to pay for such a program. If respondents did not believe that the alum treatment will work, why did they vote for the program? If respondents believed that the river and lake will naturally recover faster than what was indicated by the interviewer, why did they vote for the program? If respondents believed that the algae effects are not at all serious, why did they vote for the program? If respondents believed that it will be extremely difficult to pay the tax increase, why did they vote for the program? The logical conclusion is that these respondents agreed to pay because they knew that their agreements were in no way binding, that their votes were hypothetical.[27]

4.3.2 The referendum approach does not eliminate hypothetical bias.

The NOAA Panel's guideline for the referendum format assumes that the survey respondents will behave as they would in a real referendum. Many economists have noted that the validity of this assumption remains an open empirical question (Diamond and Hausman 1994; Fisher 1996; Green, et al. 1998; Schläpfer and Brauer 2007; Harrison 2007). Thus, it is simply conjecture to argue that the results of the Stratus

[27] See Section 5 for an analysis of how WTP varies based on these inconsistencies in logic.

study would mimic those of a real referendum because respondents in the Stratus study do not bear the consequences of their votes like they would in a real referendum. In fact, there is no compelling empirical evidence demonstrating that the hypothetical referendum format ***eliminates*** hypothetical bias.[28] On the contrary, there are many fundamental differences between a hypothetical CV question and a real referendum (Desvousges, Hudson, and Ruby 1996), indicating that hypothetical questions continue to be a problem. These differences include:

- Respondents are not required to pay the CV amount, but voters do pay for policies passed in a referendum. Therefore, the people who ultimately bear the consequences of the outcome are very different. Responsible parties bear the consequences in a CV damage assessment, and the voters bear the consequences in a referendum. Thus, in a real referendum the voters face the cost of making a mistake.
- In a CV referendum, respondents do not have to make any effort to cast their ballot, merely answer, "yes or no" to the interviewer's question. In a real referendum, voters have to make the effort to go to the polls to vote, which reflects that the issue was important enough to them to make the effort.
- In an actual referendum, the voters have a chance to confer with others whose opinions they value before casting their votes (Schläpfer 2008). The ability to air various arguments signals a very different process when compared to a hypothetical CV referendum (Horowitz 2000).
- In a CV referendum, respondents have to make up their minds on the spot during the survey interview. In a real referendum, voters have ample time to think about an issue before they cast their ballot. Voters in actual elections go to the polls knowing that they will cast a vote. When respondents first agree to participate in a hypothetical CV referendum, they do not necessarily know, at the beginning of the process, that they will be expected to "vote." Vossler and Kerkvliet (2003) identify this element of surprise for CV respondents in a hypothetical referendum as a key difference between actual votes and hypothetical ones. (See also Horowitz 2000).
- The access to information is controlled by the survey designer in a CV survey while voters have the opportunity to obtain as much or as little

[28] Hypothetical referenda may not even be good predictors of actual referenda. Diamond and Hausman (1993) argue that the referendum approach "...has no foundation in individual economic preferences" because respondents react to the amount of information they have and the context of the situation (Diamond and Hausman 1993, p. 17). They discuss an example of the inaccuracy of opinion polls on environmental issues in California ("Big Green" Proposition 128). Two opinion polls conducted by the Los Angeles Times found that 55 percent of respondents who had an opinion (84 percent of those polled) would vote in favor of the proposition. However, Proposition 128 was supported by only 36 percent of voters in the actual election. See Diamond and Hausman (1993) for additional details on how surveys were poor predictors of 36 actual election outcomes.

information as they desire in a referendum, including alternative viewpoints. In this case, the survey presented only the State's viewpoint on the matter and did not represent the viewpoint of the poultry industry, or others. It did not highlight any of the economic trade-offs that the alum treatments would impose on other farmers and ranchers, such as cattle-grazing impacts.

- In a CV referendum, respondents have to answer out loud to an interviewer. In a real referendum, voters cast their ballot in the privacy and secrecy of the voting booth. In the Stratus CV survey, not only did the respondents have to state their votes verbally to the interviewer, but in the case of approximately 30 percent of the favorable votes, other adults were present during the interview (Table D.66). The lack of privacy during a CV vote may result in an upward bias of votes "for" because the respondents may try to please the interviewer (Tourangeau, Rips, and Rasinski 2000) or may want to appear more socially responsible (Vossler, et al. 2003; Ethier, et al. 2000). Kanninen (1995) reveals that 20 percent of CV respondents may be "yea-sayers." Harrison (2007) reveals the results of a "ballot box" study where one-half of the in-person survey respondents were allowed to cast their votes on paper, without revealing them to the interviewer. The results indicate a much lower percentage of "for" votes when respondents did not have to reveal their votes verbally to the interviewer.

- In an actual referendum, the results refer to the percentage of the votes for the program, based on a pre-determined cost. In a hypothetical CV referendum, the results depend on subsequent statistical manipulations to arrive at society's purported value for the resources described. Many judgments and assumptions underlie the hypothetical survey results, while in actual elections, the election officials make no judgments or assumptions in order to determine the results.

- The multiple, fictitious dollar amounts offered in a CV question are not the actual cost of providing the public good, but are tools of the survey designer. "Estimating mean WTP requires the researcher to vary the policy's stated cost across the respondents and then calculate the implied distribution of WTP. In this case, one of two problems arise. Either the researchers must lie about the policy's costs to the respondents *or* the costs must be randomly distributed across the population" (Horowitz 2000). The evidence suggests that people anchor on those values.

- The goal of a damage assessment is to determine a specific dollar value of forgone services while the goal of a referendum is to determine whether or not some program should be adopted. The damage assessment goal requires a higher degree of precision because the absolute magnitude of the estimate is crucial.

Therefore, the argument that using a mock referendum eliminates the problem of hypothetical bias is without foundation.

Clearly, other important differences exist between the Stratus hypothetical referendum vote and those in a real referendum. For example, the first difference, and probably the most obvious, is that the interviewers showed up on their doorsteps to record their votes. Thus, respondents did not have to exert the effort to vote for the referendum as they would have to in a real referendum. Early in the survey process, respondents are told that the interviewers are on their doorsteps because the State wants to find out if people are willing to pay for a new State program. When asked whether or not they had previously been interviewed "like this" to get their opinion about whether the State should spend tax money on a new program, 98 percent of the respondents indicated that this had never happened to them before (Table D.13). From the very beginning, respondents know that this is not a normal way for a governing body to solicit public opinion. The normal ways include issuing written responses for comments, conducting public hearings and town-hall meetings, and even conducting opinion polls by telephone. Showing up on the doorstep is practically unheard of.

Moreover, the potential respondents were pursued for their opinions. There were "sorry I missed you" cards left in the door when respondents were not home. There were advance letters and refusal conversion letters in the mail. There were even refusal conversion phone calls from university professors in Maryland and California. One of the people who refused to complete the interview is quoted as saying "over and over, I'm not interested." Another refusal indicated that his wife is pregnant and they have kids and their home life is not conducive to completing a survey. The list of refusal follow-ups clearly indicates a vigorous pursuit of respondents to complete the interview (Bishop Corr 0000126). Government agencies seeking public input on tax spending matters rarely pursue public opinion with such vigor. Not only were the respondents pursued, but they were paid as well.[29] In normal public opinion matters, respondents give their opinions freely. For all of these reasons, the setting for the voting event, from the respondents' viewpoint, was not comparable to a normal referendum.

[29] Apparently, some respondents were paid $20 and some were paid $50 to try to get the more reluctant respondents to complete the survey (Appendix C of the Stratus CV report).

Another critical difference between a real referendum and the Stratus CV study is that respondents did not have a not-vote option. Certainly, in actual elections and referenda, voters may choose to not vote. Presumably, if some voters do not care about the outcome in an actual referendum, they may not bother to vote. In this study, the respondents did not have that choice. Stratus made a conscious decision to not allow a no-vote option, despite the NOAA's Panel recommendation that it be included. Specifically, the NOAA Panel included this recommendation to identify respondents who were indifferent, who needed more time or information before they could credibly vote, who preferred another mechanism, who were bored, or who wanted the survey to end. When the no-vote option is included in a referendum CV survey, the percentage of respondents who choose it ranges from 9 percent (Carson, et al. 1994) to 30 percent (Whittington, et al. 1994). As Section 4.3.1 above demonstrates, both the inclusion of and the treatment of undecided votes often changes the results of the hypothetical CV referendum.

One of the reasons that researchers do not include a no-vote option is that it will reduce the sample size on which WTP estimates are generated. While such a concern may be valid for academic study with limited funding, that restriction does not apply to the Stratus CV survey. Clearly, the Stratus researchers had ample funding to implement a survey with a large sample size.[30] Instead, Stratus cites recent research (conducted by members of its team) that it finds sufficiently compelling to disregard the NOAA Panel's recommendation. However, Harrison (2007) provides a discussion of how the results of this research "are very sensitive to how one interprets responses" (p. 94).

When disregarding the NOAA Panel's guideline, Stratus concludes that CV surveys which are "designed very carefully to use language that is clearly understandable to respondents" need not include a no-answer option (p. 3-18). Whether or not the respondents clearly understood the Stratus CV survey is highly debatable because the survey's results reveal that some respondents did not understand the survey (see Section 4.2 and Table 4.3 above). These results include

[30] However, passing the scope test appears to be an artifact of the large sample size. See Section 5 for a discussion.

confusion about the role of the litter-spreading ban, confusion about the amount of taxes that would be paid, and confusion about which rivers and lakes that tax monies and alum treatments would be applied to. The claim of having a questionnaire that is clearly understood by the respondents is an inadequate basis for rejecting the NOAA panel recommendation on the no-vote option.

Stratus also asserts that the consequentiality of the mock referendum eliminates the hypothetical bias. In essence, consequentiality refers to the realism of the survey, from the viewpoint of the respondents. A careful review of the realism aspects of this survey reveals that the Stratus CV survey is not consequential. As previously discussed, at least some respondents were impatient, distracted, inattentive, and did not take the survey seriously. The foregoing discussion has also highlighted the many ways that this mock referendum differs from a real referendum.

For consequentiality to hold, Carson and Groves (2007) also add that the respondent must believe that the government agency can compel them to pay (p. 188). However, the ability of the government agency to compel them to pay is not realistic for many respondents. Recall that the alum treatments will be funded through State income taxes. As previously described above in Section 4.3.1, many respondents who took the base version of the survey and voted for the program do not pay State income taxes. Certainly, more than one-third of the respondents knew that the State would not be able to extract payment from them. For all of the reasons discussed in this section, the Stratus CV Survey lacks consequentiality. Even if, for the sake of argument, one believed that the referendum approach eliminates hypothetical bias, this CV survey deviates too much from an actual referendum to do so.

4.3.3 The certainty question does not eliminate hypothetical bias.

One of the reasons that Dr. Bishop, one of the authors of the Stratus report, does not believe that hypothetical bias is relevant to the Stratus CV results is the use of the certainty question (Bishop undated, Bishop0001271)[31]. During the interview, shortly after the vote question, respondents are asked how sure they are of their vote. Potential responses range from “extremely sure” to “not at all sure,” with five graduated

categories for the respondent's answer. This question is known as the certainty question. The typical adjustment to WTP recodes the less certain "for" votes as "against" votes. When Stratus makes this adjustment to their results, the average WTP falls from $184.55 to $178.08 (Table G.2), which they conclude is not a significant difference.[32]

In the literature cited by Dr. Bishop (undated), he describes several empirical studies where the use of a certainty question purportedly eliminates hypothetical bias:

- Blumenscheim, et al. (1998) asked respondents whether they would (hypothetically) pay a given price for sunglasses. For those who responded affirmatively, the next question asked whether they were "definitely sure" or "probably sure" of their decision. The "probably sure" respondents were recoded to "no."
- Champ, et al. (1997) used a 10-point scale, where 10 represented "very certain" and 1 represented "very uncertain." Champ, et al. (1997) re-coded all of the "yes" responses with scores on the certainty scale other than 10 to be "no."
- Champ and Bishop (2001) used the same 10-point scale. They re-coded "yes" to "no" for all certainty scores below 8.
- Poe, et al. (2002) used the same 10-point scale. They re-coded "yes" to "no" for all certainty scores below 7.

What is noticeably different from the certainty adjustments in these four studies and the certainty adjustment in Stratus report is the extent of re-coding. In two of the four studies, only the most certain respondents' answers were not re-coded while any expression of uncertainty was re-coded. In the other two studies, "yes" responses in either the lowest 60 percent or the lowest 70 percent of the certainty scale were re-coded. In Champ and Bishop (2001), for example, the recoding resulted in almost 50 percent of the "yes" votes being re-coded to "no" votes. Only with that substantial adjustment did the hypothetical results reflect the actual results.

To be comparable to the adjustments made in these four certainty studies, Stratus should, at a minimum, re-code the "for" votes in the "moderately sure" category.

[31] Dr. Bishop wrote some thoughts or musings on hypothetical bias in CV in an undated paper.

More than 20 percent of the respondents who took the base version of the survey fall into this category, and more than 64 percent of them voted for the alum treatments (Table 6.31). Mimicking the more appropriate re-coding protocols in the studies cited above would substantially lower the WTP result.[33] Accordingly, the certainty adjustment as applied in this CV study has not eliminated hypothetical bias.[34]

4.4 The WTP estimates cannot be validated, rendering the results unreliable.

External validity[35] requires that a CV survey be capable of producing "true" WTP values for a specific commodity. External validation involves comparing values produced by CV to some objective value that has been calibrated for a high degree of accuracy. For example, Greenwich Mean Time provides a standard for evaluating the accuracy of a time clock. The time measured by a clock can be externally validated by the Greenwich atomic clock.

External validation is an important part of any scientific research because it allows the researcher to evaluate the plausibility of data, assumptions, and any model predictions. Such validation is particularly crucial for a damage assessment because claims are required to be reduced to a "sum certain." (51 *Fed. Reg.* 27751, 1986.)

Total values, comprised of both use and nonuse values, cannot be externally validated because no standard and independent measure exists for comparison. Use values can potentially be validated though revealed preference techniques (observing

[32] Had the Stratus team used the "certainty adjusted" WTP estimate for damages, the amount would be lower by almost $9 million.

[33] See Section 5 for details.

[34] Bishop (undated) also refers to the use of "cheap talk" as another mechanism for eliminating hypothetical bias. Cheap talk as used by CV practitioners and by experimental economists refers to explicit language in a CV questionnaire that defines hypothetical bias for the respondent, emphasizes that the vote in this survey is hypothetical, but asks the respondent to vote as if it were real. There are various gradations of cheap talk, with "heavy" cheap talk being more explicit in terms of defining hypothetical bias and emphasizing the hypothetical nature of the survey. According to Dr. Bishop, "light" cheap talk produces mixed results with respect to hypothetical bias but "heavy" cheap talk "solves the problem of hypothetical bias. Dr. Bishop notes that Stratus did not employ a heavy cheap talk strategy because "[a] contingent valuation survey cannot be consequential if it states in no uncertain terms that the whole exercise is hypothetical" (p. 17). This presents a conundrum. Heavy cheap talk allegedly eliminates hypothetical bias but it also eliminates consequentiality, the presence of which also purportedly eliminates hypothetical bias.

[35] External validity is also commonly referred to as *criterion validity*.

recreator behaviors). In contrast, nonuse values are not associated with an observable behavior and cannot be measured using a revealed preference technique or market prices. Because nonuse values are a component of total value, a revealed-preference technique cannot be used to determine total value either. CV total values cannot be externally validated because of the absence of alternative estimation methods, such as market-place transactions or revealed-preference techniques, which can directly confirm or refute total values.

Other economists recognize the lack of a "true" value to compare with contingent-valuation total values. For example, Freeman (1993) states that "[i]deally, one would like to assess the validity of a hypothetical value by comparing it with the true value. But the true value is usually not known, so this option is not available" (p. 176). Similarly, Smith (1986) states that "the only standard available from current research is itself an estimate of the unknown 'true' value of an individual's valuation" (p. 174).

4.4.1 The Stratus scope test is not meaningful

A scope test is an essential part of a CV study. The test consists of administering two versions of a survey questionnaire to two samples of respondents. The questionnaire versions are identical in everything but the magnitude of the environmental injury. In other words, the injury described in the scope version ("scope survey") will be smaller in magnitude than the injury described in the main version ("base survey"). Respondents are randomly assigned to one version so that the respondents groups across the two versions are as identical as practical.

CV studies suffer from an "embedding" effect: it has been observed that WTP to mitigate an environmental problem affecting a large area is not very different from WTP to address the same problem in a small subpart of that area (Kahneman and Knetsch 1992). It has also been observed that WTP differs little based on the amount of a particular wildlife species that will be conserved (Desvousges, et al. 1993; Arrow, et al. 1993). These results are at odds with basic economic principles that dictate WTP to resolve a larger scale problem should be greater than for a smaller scale problem.

The best explanation for this phenomenon is that respondents did not actually report the economic value of the good they were asked about, but rather were deriving "moral satisfaction" (Kahneman and Knetsch 1992) from being able to contribute to the mitigation of the problem and being known to do so. As this "moral satisfaction" (or "warm glow") remains relatively consistent despite the scale of the environmental problem, the reported WTP will also be similar.

The problem presented by these "moral satisfaction" or "warm glow" findings led to the following statement in the NOAA Blue Ribbon Panel Report (Arrow, et al. 1993, p. 37):

> "Specifically, if a CV survey suffered from any of the following maladies, we would judge its findings 'unreliable': [...] -- Inadequate responsiveness to the scope of the environmental insult."

As a result, determining whether the responses were unreliable because they are inadequately responsive to changes in the scope of an environmental problem has become a requirement for properly conducted CV studies used in NRDA cases. The members of the NOAA panel (Arrow, et al. 1994) later clarified what they meant by "inadequate:"

> "The report of the NOAA panel calls for survey results that are 'adequately' responsive to the scope of the environmental insult... Had the panel thought that something as straightforward as statistical measurability were the proper way to define sensitivity, then we would (or should) have opted for language to that effect. A better word than 'adequate' would have been 'plausible': A survey instrument is judged unreliable if it yields estimates which are implausibly unresponsive to the scope of the insult. This, of course, is a judgment call, and cannot be tested in a context-free manner, as would be the case if the proposed scope test were implemented.
>
> These two definitions will not generally yield the same conclusions.
>
> There will be settings in which estimates made with plentiful observations are 'statistically' sensitive to the scope but at the same time are 'implausible' [sic] insensitive. Also, if the sample size is small and the scope difference minor, the estimates may be 'statistically' insensitive to the scope, yet 'plausibly' sensitive.

> The fundamental problem with any purely statistical sensitivity is that it depends (foolishly) on the sample size.
>
> In small samples, no effects are 'statistically significant.' In large samples, everything is 'statistically significant.' What this means is that the proposed scope test can probably be passed if the trustees are willing to pay a high enough cost. But the willingness to bear this cost has no obvious implications for the 'reliability' of the results."

The authors of the Stratus Report re-characterize the NOAA Guidelines by stating: "The Panel was referring to the expectation, based on economic theory, that WTP to achieve a ***larger environmental improvement*** should be larger than WTP to avoid a smaller one [emphasis added]."[36] This unusual interpretation leads the authors to simultaneously distinguish the base scenario (designed to be valued) from the scope scenario (which is used to test it), across more than one dimension: both geographic scope and effectiveness of the proposed treatment. A proper scope test should distinguish only one dimension: "either in a quality or quantity sense" (Carson, Flores, and Meade 2001, p. 181).

The distinctions between the base survey and scope survey scenarios in the Stratus Report illustrate the multi-dimensional differences in their coverage:[37]

a. In the scope scenario, the target of the alum treatments is restricted to the lake and does not include the river. The scope version of the questionnaire indicates that the river would recover naturally in 10 years. There is no mention of trucks to spread alum on the land and dispensers in the river.

b. In the scope scenario, phosphorus levels in the river would return to 1960 conditions 10 years following the ban without any treatment, whereas, in the base scenario, those levels recover in 50 years without any treatment, or 10 years with the alum treatment.

c. The base survey states that the lake will recover in 20 years with alum treatments, whereas the scope survey lengthens that period to 50 years. This enlargement of 30 years required for recovery runs directly contrary to the purpose of the Scope Survey, which is to ***reduce*** the scope of the problem and determine whether respondents scale their WTP accordingly.

[36] Emphasis Added, Stratus Report, Vol. I, p. 3-17.
[37] The differences between the texts of the two surveys are highlighted in Appendix B.

Although the injury described in the scope survey is smaller than the base survey in terms of geography (a and b above), it is larger in its persistence over time (c above). Further, contrary to the NOAA guidelines that call for a "high standard of richness in context to achieve a realistic background,"[38] the Stratus report does not tell respondents the exact size of the area that is described in each of the questionnaires. Item (c) incorporates differences in both persistence of the articulated problem and the effectiveness of treatment with alum. Respondents likely perceive 50 years with alum treatment as a comparatively small improvement when compared with the 20 years described in the base instrument for recovery with alum. It is not clear why there should be a difference in the effectiveness of alum treatments between the two scenarios.

By diverging from the base scenario along multiple dimensions it is virtually assured that respondents will view the two scopes differently but not in a manner that can be used to test whether WTP is appropriately sensitive. The differences in valuation are not just due to a "smaller injury," but also to perceived differences in the effectiveness of treatment. Accordingly, this scope test cannot be used to affirm the WTP contained in the Stratus Report.

The literature on CV recognizes that problems in survey design can also result in failure to pass a scope test (Carson, Flores, and Meade 2001, p. 181). Among these problems are "…questions where the underlying metric on which respondents perceive the *larger* good is different from that on which respondents perceived the *smaller* good, and…differences in the perceived probability of the different goods actually being provided" (Carson, Flores, and Meade 2001, p. 200). In other words, if respondents think the proposed environmental solution in the scope survey is more or less likely to work than the one described in the base survey, this will lead to differences in their responses. Similarly, where respondents view two different environmental resources (e.g., a river and a lake) as having different uses or values, this may result in a different WTP for restoration. In either of these examples, the differences perceived by respondents between the two survey versions will influence the results, which can no longer properly be used to satisfy the scope test.

[38] NOAA report, p. 28

A number of studies have criticized precisely this type of error or noted the need to isolate a single dimension of relative injury in the design of any scope test. Among these are Smith and Osborne (1996), in which a meta-analysis was conducted to answer the scope question, using different CV studies that measure WTP for visibility in parks. The study compared sensitivity to scope only on one dimension – relative visibility. Similarly, Carson (1997) noted: "The oil spill experiment is marred by the fact that the larger good invoked a different (and lower) probability of success of preventing a large spill than had been used in the second treatment, thereby providing a significant confounding factor." This is precisely the type of confounding factor present in the Stratus surveys.

The authors of the Stratus Report mention that 58.4% of respondents to the base survey voted "for" the proposed cleanup program.[39] They fail to report the same measure for the scope survey: which is 42.5% in favor. This result may be driven by the fact that respondents had less faith in the effectiveness of the remediation scenario described in the scope survey and not just the smaller magnitude of the described injury. The authors provide no analysis of how to unravel these two confounding influences.

Average WTP derived from the base survey is $184.55, leading to an aggregate WTP of $249,673,635.[40] By contrast, average WTP under the scope survey is $138.51, leading to aggregate WTP of $187,387,131.[41] This reduction in WTP is only 25%, a surprisingly small change in response to what is supposed to be a significantly smaller environmental damage.[42] To put this into perspective, the difference between the base and scope WTP in a 1994 CV study, whose authors include Stratus team members Hanemann and Krosnick, was nearly twice that of the Stratus study at

[39] Stratus Report, Vol. I, p. 6-2.
[40] $184.55 x 1,352,878=$249,673,635.
[41] $138.51 x 1,352,878= $187,387,131.
[42] The scope WTP is about 75 percent of the base WTP. If one believed the WTP results from the Status CV study, then Oklahoma residents are willing to pay $138 dollars to accelerate restoration of Lake Tenkiller from 60 years to 50 years. But they are willing to pay less than $46 ($184 minus $138) to accelerate restoration in the Illinois River by 40 years, and achieve faster recovery of the lake, relative to the scope version. This illogical conclusion is the result of inappropriate survey design.

47%.[43] The Court ultimately rejected that study of damages from PCB and DDT contamination, apparently because "the descriptions of alleged 'injuries' to fish and birds used in the survey were unsupported by the trustees' own evidence and experts."[44] As previously discussed, the same problem exists with the Stratus survey due to misleading, incomplete and arguably inaccurate factual statements about algae and its impact on fish populations as well as the benign effects of alum treatments.

Information about the reasons respondents provided for voting "for" or "against" the referendum question also fail to assist in resolving this confusion between the base and scope scenarios. Tables 6.28 and 6.30 of the Stratus report list reasons respondents supplied for voting for or against the referendum question in the scope survey. The authors assert that "the reasons… for voting 'for' and 'against' the program… closely resemble the reasons given by the base respondents" (page 6-31). However, if we closely compare the corresponding tables for the base survey (Table 6.2 and Table 6.3), a number of important differences are apparent. All four tables (two pertaining to the base survey and two pertaining to the scope survey) bear a notation that the percentages they list may not total to 100 percent, because respondents could have supplied more than one reason for their votes. However, the base survey responses total 122.2 percent (reasons for "Yes" votes) and 132.8 percent (reasons for "No" votes), while the scope survey reasons total precisely 100 percent ("Yes" reasons) and 100.6 percent ("No" reasons). This suggests that respondents were differently encouraged to complete and fully respond to the two surveys.

Leaving aside this problem of number of reasons supplied per respondent, those voting "Yes" in the base survey appear to have been more convinced of the efficacy of the proposed restoration than were those who voted "Yes" in the scope survey referendum. Nearly 40 percent of those who voted "Yes" in the former group said the program would help the area around the lake, while only 34.1 percent of those in the latter group offered the same explanation. Similarly, 20.6 percent of those voting

[43] Carson, Richard T., W. Michael Hanemann, Raymond J. Kopp, Jon A. Krosnick, Robert C. Mitchell, Stanley Presser, Paul A. Ruud, and V. Kerry Smith. 1994. "Prospective Interim Lost Use Value Due to PCB and DDT Contamination in the Southern California Bight" (Report to National Oceanic and Atmospheric Administration): p. 253.

[44] Court Rejects Contingent Valuation Study in Montrose Case. (June 2000). Sidley & Austin Environmental Advisory.

yes in the base survey said the program would speed up recovery, while only 15.9 percent of those voting yes to the scope survey provided this same response. These differences between favorable votes in the base and scope survey are detailed in Table 4.5.

Table 4.5: Comparison of Base and Scope Reasons for Voting "For"

Base Survey Responses	%	Scope Survey Responses	%
1. Program will help area around river and lake	39.6%	1. Program will help area around river and lake	34.1%
2. Program will benefit others	22.2%	2. Program will benefit others	16.6%
3. Program will speed up the recovery	20.6%	3. Program will speed up the recovery	15.9%
4. Program will bring lake back to earlier state	11.5%	4. Other	7.8%
5. Other	9.0%	5. Program will bring lake back to earlier state	6.8%
6. Program reduces risk to human health	4.7%	6. Program reduces risk to human health	5.1%
7. Respondent would benefit from program	4.2%	7. Program will protect environment in general	4.7%
8. Program will protect environment in general	4.2%	8. Respondent would benefit from program	3.7%
9. Respondent is concerned about environment	2.6%	9. Respondent is concerned about environment	2.7%
10. Program has other benefits than cleaning water	2.3%	10. Society is responsible for fixing problem	1.0%
11. Society is responsible for fixing problem	0.9%	11. Program has other benefits than cleaning water	1.0%
12. Others in household concerned about environment	0.2%	12. Others in household concerned about environment	0.3%
13. Don't know/ Doesn't remember	0.2%	13. Blank response	0.3%
Total	122.2%	Total	100%

These differences are more pronounced when comparing the reasons for voting "No" on the referendum as described in Tables 6.3 and 6.28 of the Stratus Report. The most common reason for voting "No" in the base survey was that the "cost is too high for respondent household" (18 percent) but this was only the fifth most common response for those voting the same way on the scope survey (8.3 percent). Conversely, for those voting "No" to the scope survey, the fourth most common explanation was that the "benefits are not worth the cost" (8.5 percent) but this was one of the least frequent explanations offered by those who voted "No" to the base survey

(only 2.4 percent, the 17th most commonly offered response). This once again suggests that, by mingling different timelines and natural recovery periods with the difference in geographic dimension of the two surveys, Stratus left respondents with two very different perceptions about the cost effectiveness of the two programs. This violates the very purpose of a scope survey. Consistent with this observation, by far the most common reason for voting "No" to the scope survey was that the "program does not do enough" (18.5 percent), while this same reason (together with "users should pay") ranked last in explaining "No" votes to the base survey (1.2 percent). This dramatic difference validates our concern that the scope survey proposed a less effective restoration program when compared with the natural recovery alternative than did the base survey. The comparison of all reasons offered for voting "No" to the two surveys appears in Table 4.6.

Table 4.6: Comparison of Base and Scope Reasons for Voting "Against" the Tax

Base Survey Responses	%	Scope Survey Responses	%
1. Cost is too high for respondent/household	18.0%	1. Program does not do enough	18.5%
2. Other issues are more important	15.1%	2. Let nature solve the problem	12.0%
3. Let nature solve the problem	13.2%	3. Other issues are more important	11.8%
4. Does not want to pay more taxes	9.6%	4. Benefits are not worth cost	8.5%
5. Cost of program is too high	9.4%	5. Cost is too high for Respondent/Household	8.3%
6. Other	8.6%	6. Cost of program is too high	6.5%
7. Not sure if the program would work	7.4%	7. Other	5.3%
8. Cost too high for others	6.5%	8. Does not care about the problem	4.8%
9. Program only helps a few rivers and/or lakes	6.2%	9. Problem could/should be solved other ways	3.8%
10. Treatments could have unknown bad effects	6.0%	10. Program only helps a few rivers and/or lakes	3.5%
11. Does not care about the problem	5.8%	11. Does not want to pay more taxes	3.3%
12. Does not like way payment would be collected	5.0%	12. Polluters should pay	2.8%
13. Problem could/should be solved other ways	4.6%	13. Cost too high for others	2.5%
14. Polluters should pay	3.6%	14. Treatments could have unknown bad effects	2.3%
15. Someone else should pay	2.9%	15. Not sure if the program would work	1.8%

Base Survey Responses	%	Scope Survey Responses	%
16. Money might be used for other purposes	2.4%	16. Someone else should pay	1.5%
17. Benefits are not worth cost	2.4%	17. Does not like way payment would be collected	1.0%
18. People not near lake won't want to pay	1.9%	18. Money might be used for other purposes	1.0%
19. Does not trust government	1.4%	19. Users should pay	0.5%
20. Program does not do enough	1.2%	20. Program would be done for other rivers/lakes	0.3%
21. Users should pay	1.2%	21. Does not trust government	0.3%
22. Don't know/ Doesn't remember	0.2%	22. Don't know/ Doesn't remember	0.3%
23. Refused	0.2%		
Total	**132.8%**	**Total**	**100.6%**

Finally, the sample size used for the base survey is approximately twice that of the scope version. The larger sample size inherently produces a smaller standard error, which translates to a smaller confidence interval for the results of the base survey. Had the base survey sample been as small as the scope survey sample, the results of the base survey would likely have had a larger standard error. Because WTP estimated from the two surveys is separated by only $3 (the difference between the upper-end of the scope WTP and the lower-end of the base WTP), a larger standard error would likely have resulted in overlapping confidence intervals. With overlapping confidence intervals, Stratus would not have been able to declare that the WTP values for the base version and the scope version are significantly different. This means that the CV would have failed the essential scope test.

A large enough sample size can make any difference statistically significant. To quantify the effect of sample size upon statistical significance in Stratus' WTP estimates, we employ a bootstrap technique for the base dataset. The goal of a bootstrap is to generate a distribution of estimates based on different samples from the base data. Instead of calculating WTP from the entire base dataset of 1,093 observations, we calculate WTP from 544 observations randomly drawn from that base data. Since the scope data consist of 544 observations, this procedure generates a hypothetical dataset of bids and responses to the base scenario that is the same size as the scope data. Not surprisingly, using comparable sample sizes alters Stratus'

claimed statistical significance of differences in WTP between the base and scope surveys.

Applying the traditional bootstrap procedure, the base survey yields a 95% confidence interval of WTP between $162.32 and $206.77, a range $6.79 wider than that derived using the entire base dataset ($165.72 to $203.38).[45] The new base WTP using this smaller sample size thus overlaps WTP estimated from the scope survey using a traditional bootstrap, which was $112.69 to $164.32, also measured at a 95% confidence interval using 544 observations.[46] This overlap of confidence intervals indicates that Stratus' selection of sample size may be driving the statistical significance it claims to identify between the base and scope WTP.

Table 4.7 reports a simulation to help illustrate this point. The first two rows contain the results reported by Stratus. Notice that the standard error (se) for the scope version is almost 30 percent higher than is the standard error for the base version. To simulate how a sample size equal to that of the scope version would change the confidence interval of the base version, the third row shows the estimated se and confidence interval for the re-estimated base survey using a traditional bootstrap. In order to compare the base survey with the smaller sample size to the scope survey, the se and confidence interval estimated using a traditional bootstrap is shown in the fourth row. This re-estimation shows that the confidence intervals now overlap. The lower end of the base version is $162.32, which is smaller than the upper end of the scope version, $164.32 and thus they overlap by $2. With a larger standard error, the WTP results are no longer statistically different and the scope survey cannot be used to validate the results of the base survey as required by professional standards.

[45] A jackknife bootstrap cannot be used to estimate the base survey WTP using a smaller random sample because a jackknife bootstrap would proceed by repeatedly calculating WTP based on the given dataset, leaving out only one observation at a time. This non-random procedure cannot exclude more or less than one observation at a time, making it impossible to simulate a smaller dataset.

[46] Scope WTP was estimated using a traditional bootstrap because the confidence intervals generated by the traditional bootstrap differ from the 95% confidence intervals generated by the jackknife. This is because the jackknife bootsrap incorporates the survey weights when it draws observations for the bootstrap procedure.

Table 4.7: Impact of Different Samples Sizes

Survey	Sample Size	Mean	se	Confidence Interval	
Base, Stratus (Jackknife Bootstrap)	**Base** (1,093)	$184.55	9.61	$165.72	$203.38
Scope, Stratus (Jackknife Bootstrap)	**Scope** (544)	$138.51	12.25	$114.50	$162.51
Base Version (Traditional bootstrap)	**Scope** (544)	$184.55	11.34	$162.32	$206.77
Scope (Traditional Bootstrap)	**Scope** (544)	$138.51	11.34	$112.69	$164.32

Some members of the NOAA Panel explicitly warned of this problem, cautioning against the use of sample size as a means to pass the scope test. "In large samples, everything is 'statistically significant.' What this means is that the proposed scope test can probably be passed if the trustees are willing to pay a high enough cost. But the willingness to bear this cost has no obvious implications for the 'reliability' of the results" (Arrow, et al. 1994). In other words, the difference that Stratus reports between the results of the base and scope surveys may largely be an artifact of sample size and do not provide the necessary indicia of reliability for the contingent valuation. This problem, compounded with errors in survey design that result in measuring fundamentally different assets, mean that the Stratus conclusions fail to cross the professionally imposed hurdle for scope validation, and therefore, the results of the Stratus study cannot be viewed as either valid or reliable.

4.4.2 The "construct validity" model does not validate the WTP estimate.

In an attempt to demonstrate validity in the WTP results, Stratus develops and reports a "construct validity" model (Table 6.26). This model explains the probability that a respondent voted for the alum treatments as a function of several of the other responses in the questionnaire. For example, the model indicates that the higher the bid amount, the less likely that the respondent would vote for the program. Similarly, the model indicates that if a respondent believed that the tax money would be used to clean up other rivers and lakes beyond the Illinois River watershed, then that respondent was more likely to vote for the program. If a respondent believed that natural recovery would occur faster than the interviewer said it would, then he/she was less likely to vote for the program. If respondents believed that the tax amount would

be higher than that indicated by the interviewer, then they were less likely to vote for the program.

Although this model demonstrates that there are patterns in the hypothetical votes and the other answers in the survey, it does not adequately explain the hypothetical votes. Based on the pseudo-R^2 of 0.31 reported by Stratus, the model only accounts one-third of the variation of the base version votes. Although the model indicates that respondents who believed that natural recovery would occur faster were less likely to vote for the program, some of them still voted for the program. Although the model indicates that respondents who believed that the tax amount would be higher were less likely to vote for the program, some of them still voted for the program. Thus, the identification of these statistical relationships is not compelling. The model does nothing more than identify some weak patterns in the hypothetical data. Nor does any explanation of hypothetical data imply that these results would reflect the actual behaviors of the respondents.

Furthermore, the sensitivity assessment conducted by Stratus also fails to demonstrate that the CV results are valid. In general, the sensitivity analysis addresses two topics: how certain the respondent was of his/her vote and the extent that the respondent accepted the scenario described by the interviewer. To implement its sensitivity analysis, Stratus alters the respondents' answers to the vote certainty question (question 24) and the scenario acceptance questions (questions 27, 29-33).[47] They re-code the data so that all respondents were at least moderately certain of their votes and so that all respondents accepted the various elements of the scenario described. When the data are altered, the model does only a slightly better job of correctly predicting the hypothetical votes. With a pseudo--R^2 of 0.33, the adjusted model still only for accounts for about one-third of the variation of the hypothetical votes. Even when Stratus assumes that the respondents were more certain of their votes and were completely accepting of the survey scenario, the model still fails to sufficiently explain the underlying motivations for the hypothetical votes. Moreover, it

[47] Incidentally, the sensitivity analysis excludes one question that also reveals the extent to which respondents accepted the scenario. Question 28 asked respondents whether they believed that phosphorus had caused the changes described, which reflects another dimension of scenario acceptance.

fails to provide any convincing evidence that either the responses or the model conforms to what people would do if they were actually required to pay the stated increases in taxes.

A more logical and conservative sensitivity analysis would have changed respondents' hypothetical votes to be consistent with the certainty of their votes and their lack of scenario acceptance. For example, if a respondent voted for the program but then said that he was not at all certain of his vote, the sensitivity analysis could have changed his hypothetical vote to "against." If a respondent voted for the program but then said that she did not think that the alum treatments would work well at all, the sensitivity analysis could have changed her hypothetical vote to "against." As discussed below, implementing these changes results in dramatically different, and lower, WTP estimates, thereby further demonstrating the lack of reliability of the Stratus CV study.[48] For all of the reasons described in this section, the "construct validity" models fail to validate the WTP responses.

4.4.3 The error rate for this CV survey cannot be known.

One of the critical characteristics of reliability is whether the methodology has a known error rate. The error rate refers to the difference in the actual result and the statistical estimate. In this specific application, the error rate represents the difference between respondents' "true" values for a faster recovery of the Illinois River System and Tenkiller Lake and the results of the Stratus CV survey. Given that the total values addressed in the Stratus CV report include nonuse values, there is no way of knowing respondents' true values. Thus, an error rate for this survey cannot be estimated.

Moreover, the specific attempts made by Stratus at establishing validity for this survey have not been successful. The Stratus scope test is not meaningful. Moreover, the results do not conform to established economic principles, as indicated in the elasticity evaluation discussed below. Respondents admitted that they were valuing a much larger commodity when they voted, and the Stratus analysis failed to account for them. The Stratus validity model itself fails to sufficiently explain the determinants of the hypothetical votes. The respondents did not seriously consider their budget

constraints when voting. For all of these reasons, an error rate for this CV survey cannot be estimated, further demonstrating the lack of reliability of the Stratus CV survey.

4.5 WTP values are artifacts of the bid design.

The WTP estimation approach used by Stratus relies on the proportion of respondents who voted "for" the restoration program at various tax payments, known as bid levels. The mean WTP represents a weighted average of the bids, with weights derived from the proportion of the sample voting for the program at each bid. Given this approach, the bid structure plays an important role in the calculation of WTP because the bids selected for the survey design affect the magnitude of WTP. This section describes the WTP's sensitivity to the bid structure, especially to the highest bid offered. The arbitrary nature of the selected bids, which are selected by the plaintiffs' experts, undermines the reliability of the estimated mean WTP.

Comparison of the results of the Stratus CV Study to other studies with different scopes of injury illustrates this point. Table 4.8 describes the key features of three other CV studies conducted as part of a damage assessment. These three studies are the *Exxon Valdez* study (Carson, et al. 1992), the Southern California Bight study (Carson, et al. 1994), and the California Oil Spill Study (COS) (Carson, et al. 1996). The latter two studies were developed to purportedly comply with the NOAA panel guidelines. These comparisons provide some evidence that CV values are artifacts of the bid design. Table 4.8 highlights the major features of each study, including the commodity, injury description, payment vehicle, location, and sampling frame. All three studies used a one-time tax-payment vehicle, minimizing any effect from payment terms across the studies.

[48] See Section 5 for further analysis.

TABLE 4.8: COMPARISON OF MAJOR DESIGN FEATURES IN THE *Exxon Valdez*, Southern California Bight, AND California Oil Spill STUDIES

STUDY	*Exxon Valdez*	Southern California Bight	California Oil Spill
Commodity	• Program to prevent an *Exxon Valdez*-type spill along the South Central Alaskan coast sometime over the next ten years	• Program to reduce the recovery period from 50 years to 5 years for four species affected by sediments contaminated with DDT and PCBs along the South Coast of California through covering of contaminated sediments.	• Program to prevent harm from oil spills along California's Central Coast over the next ten years
Injury	• 1,000 miles of shoreline oiled (few years to recover) • 22,600 birds found dead—estimated total bird deaths of 75,000 to 150,000 (three to five years to recover) • 580 otters and 100 seals killed (couple of years to recover)	• Each of the four species' (two fish and two endangered bird species) reproductive abilities have been affected by DDT and PCB contamination (five years to recover)	• Many small animals and plants along ten miles of coastline (five years to recover) • 12,000 birds killed and 1,000 injured (ten years to recover)
Payment Vehicle	• One-time increase in federal income taxes; money sent to Prince William Sound Fund	• One-time increase in state income tax; money would go to a special fund to cover the contaminated sediment	• One-time increase in state income taxes; money goes toward setting up response centers
Sampling Frame	• U.S. residents	• English-speaking households in California	• English-speaking households in California

The *Exxon Valdez* study focused on a larger oil spill with more extensive injury to mammals and birds. Geographically, it covered 1,000 miles of shoreline. The restoration program would prevent similar oil spills for the next 10 years. The Southern California Bight study described injuries to two endangered species of birds along the southern coast of California. It shares a similarity with the Stratus CV survey in that the restoration program would accelerate recovery of the resources from 50 years to 5 years. The COS study described injuries to common bird species along the central coast of California. Its restoration program would prevent harm from oil spills for the next 10 years.

The Stratus CV survey described different restoration conditions in the base and scope versions. In the base version, the survey described that algae impacted about

60 miles of the Illinois River and Tenkiller Lake, which is approximately 28 miles long. The survey noted that the Illinois River and several of its tributaries are designated as scenic rivers. The restoration program in the base survey would accelerate recovery from 50 years to 10 years for the river and from 60 years to 20 years for the lake. In the scope version, the restoration program would only affect the lake and would accelerate recovery by only 10 years (from 60 years to 50 years).

Table 4.9 compares the mean WTP values from the studies.[49] For the Stratus CV study, both the mean values for the scope and base version are included. What should drive the differences in per household WTP values is the relative size of the injury described. Thus, intuition suggests that the WTP value from the *Exxon Valdez* oil spill study, which arguably describes the most extensive injury of those in the table, should have the highest WTP value. However, the *Exxon Valdez* study has the lowest WTP value because it has the lowest bid structure. The top bid included in the *Exxon Valdez* study was $120.

TABLE 4.9: Comparison of Mean Bids for CV Surveys

	Exxon Valdez	**Southern California Bight**	**California Oil Spill**	**Stratus CV Survey**
Number of Initial Bids Offered	4	5	5	6
Bid Range	($10 – $120)	($10 – $215)	($5 – $220)	($10 – $405)
Mean*	$84.30	$108.45	$150.02	$138.51 (scope) $184.55 (base)

* Means for the first three are calculated using the Turnbull estimator. Mean from *Exxon Valdez* comes from Table E2-7 (Table 2) in Carson, et al. (1996) report. Southern California Bight mean comes from p. 195 in Carson, et al. (1994) report (Table 9.4). COS mean comes from p. 6-7 (Table 6.3) in Carson et al. (1996) report. Mean values are expressed in 2008 dollars for all four studies.

[49] Although the *Valdez*, California Bight, and COS studies used a different nonparametric estimator than does the Stratus Study, the general conclusions still hold. Section 5 of this report re-analyzes the Stratus data using the same nonparametric estimator used by the other studies. It also demonstrates that the Stratus pretest data, which reflect different (and lower) bid designs, confirms the impact on WTP from the bid design.

In terms of the Stratus CV survey, its highest bid is almost twice as large as the highest bids California Bight and the COS studies and almost four times larger than the highest bid in the *Exxon Valdez* survey. The amount of the highest bid is what dominates the WTP values, not the differences in the injury described. The Stratus WTP result is more than two times the WTP from the Exxon Valdez survey, which is remarkable considering the differences in the described injuries. While it is possible that some of the difference is attributable to preferences changing over time and preferences differing among the households surveyed, the difference is too large to be explained by these other factors.

Suppose for the sake of argument, we use the interpretation of the scope design presented by Stratus, and ignore the differences in restoration efficiencies discussed in the previous section. We then compare the mean WTP for the scope version relative to the base version in the Stratus CV study. This comparison highlights the inappropriate influence of the bid design based on the samples from the same population of respondents. According to the Stratus design, the scope version focuses only on the lake, not the river. The lake is about half the size of the river in terms of miles. Moreover, according to the Stratus design, the acceleration of restoration for the lake is markedly different in the two versions. According to the Stratus design, the scope version, the acceleration of restoration is much smaller. Specifically, the acceleration is only 10 years of difference, from 60 years to 50 years. If people ignored the confounding effects we described in the previous section, economic principles predict that the mean WTP would be smaller in the scope version.[50] However, comparison to the base version reveals the influence of the bid design. The mean WTP for the scope version is 75 percent of the mean WTP for the base version. If one were to believe the WTP results from the Status CV study, then Oklahoma residents are willing to pay $138 dollars to accelerate restoration of Tenkiller Lake from 50 years to 60 years, but are willing to pay less than $46 ($184 minus $138) to accelerate restoration in the Illinois River by 40 years. (Because the base version also accelerates the recovery of the lake at a rate faster than that described in the scope version, presumably some of the residual $46 would reflect an even faster recovery for the

[50] NOAA has warned against selecting two levels that are so vastly different that passing a scope test is a foregone conclusion (59 *Fed. Reg.* 1146).